

# Ancestral Graph Markov Models\*

Thomas Richardson  
University of Washington, USA  
tsr@stat.washington.edu

Peter Spirtes  
Carnegie-Mellon University, USA  
ps7z@andrew.cmu.edu

15 April 2002

## Abstract

This paper introduces a class of graphical independence models that is closed under marginalization and conditioning but that contains all DAG independence models. This class of graphs, called maximal ancestral graphs, has two attractive features: there is at most one edge between each pair of vertices; every missing edge corresponds to an independence relation. These features lead to a simple parametrization of the corresponding set of distributions in the Gaussian case.

**Key words:** Directed acyclic graph, DAG, ancestral graph, marginalizing and conditioning, m-separation, path diagram, summary graph, MC-graph, latent variable, data-generating process.

## 1 Introduction

The purpose of this paper is to develop a class of graphical Markov models that is closed under marginalizing and conditioning, and to describe a parametrization of this class in the Gaussian case.

---

\*Technical Report No. 375, Department of Statistics, University of Washington, Box 354322, Seattle WA 98195, USA; also No. 370, Department of Statistics, University of Warwick, Coventry CV4 7AL, UK.

Report Documentation Page				Form Approved OMB No. 0704-0188	
Public reporting burden for the collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to a penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number.					
1. REPORT DATE <b>15 APR 2002</b>		2. REPORT TYPE		3. DATES COVERED <b>00-00-2002 to 00-00-2002</b>	
4. TITLE AND SUBTITLE <b>Ancestral Graph Markov Models</b>				5a. CONTRACT NUMBER	
				5b. GRANT NUMBER	
				5c. PROGRAM ELEMENT NUMBER	
6. AUTHOR(S)				5d. PROJECT NUMBER	
				5e. TASK NUMBER	
				5f. WORK UNIT NUMBER	
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) <b>University of Washington, Department of Statistics, Box 354322, Seattle, WA, 98195</b>				8. PERFORMING ORGANIZATION REPORT NUMBER	
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)				10. SPONSOR/MONITOR'S ACRONYM(S)	
				11. SPONSOR/MONITOR'S REPORT NUMBER(S)	
12. DISTRIBUTION/AVAILABILITY STATEMENT <b>Approved for public release; distribution unlimited</b>					
13. SUPPLEMENTARY NOTES					
14. ABSTRACT <b>This paper introduces a class of graphical independence models that is closed under marginalization and conditioning but that contains all DAG independence models. This class of graphs, called maximal ancestral graphs, has two attractive features: there is at most one edge between each pair of vertices; every missing edge corresponds to an independence relation. These features lead to a simple parametrization of the corresponding set of distributions in the Gaussian case.</b>					
15. SUBJECT TERMS					
16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT <b>Same as Report (SAR)</b>	18. NUMBER OF PAGES <b>87</b>	19a. NAME OF RESPONSIBLE PERSON
a. REPORT <b>unclassified</b>	b. ABSTRACT <b>unclassified</b>	c. THIS PAGE <b>unclassified</b>			

A graphical Markov model uses a graph, consisting of vertices and edges to represent conditional independence relations holding among a set of variables (Lauritzen, 1979, Darroch et al., 1980). Three basic classes of graphs have been used: undirected graphs (UGs), directed acyclic graphs (DAGs), and chain graphs which are a generalization of the first two. (See Lauritzen, 1996; Whittaker, 1990; Edwards, 1995.)

The associated statistical models have many desirable properties: they are identified; the models are curved exponential families, with a well-defined dimension; methods for fitting these models exist; unique maximum likelihood estimates exist.

All of these properties are common to classes of models based on DAGs and UGs. However, as we will now describe there is a fundamental difference between these two classes.

Markov models based on UGs are closed under marginalization in the following sense: if an undirected graph represents the conditional independencies holding in a distribution then there is an undirected graph that represents the conditional independencies holding in any marginal of the distribution. For example consider the graph  $\mathcal{U}_1$  in Figure 1(i) which represents a first-order Markov chain. If we suppose that  $y_2$  is not observed, then it is self-evident that the conditional independence,  $y_1 \perp\!\!\!\perp y_4 \mid y_3$ , which is implied by  $\mathcal{U}_1$  is represented by the undirected graph  $\mathcal{U}_2$  in Figure 1(ii), which does not include  $y_2$ . In addition,  $\mathcal{U}_2$  does not imply any additional independence relations that are not also implied by  $\mathcal{U}_1$ .



Figure 1: (i) an undirected graph  $\mathcal{U}_1$ ; (ii) an undirected graph  $\mathcal{U}_2$  representing the conditional independence structure induced on  $\{y_1, y_3, y_4\}$  by  $\mathcal{U}_1$  after marginalizing  $y_2$ .

By contrast Markov models based on DAGs are not closed in this way. Consider the DAG,  $\mathcal{D}_1$ , shown in Figure 2(i). This DAG implies the following independence relations:

$$t_1 \perp\!\!\!\perp \{t_2, y_2\} \quad t_2 \perp\!\!\!\perp \{t_1, y_1\} \quad (\ddagger)$$

DAG  $\mathcal{D}_1$  could be used to represent two successive experiments where:

- $t_1$  and  $t_2$  are two completely randomized treatments, and hence there are no edges that point towards either of these variables;
- $y_1$  and  $y_2$  represent two outcomes of interest;
- $h_0$  is the underlying health status of the patient;
- the first treatment has no effect on the second outcome hence there is no edge  $t_1 \rightarrow y_2$ .

There is no DAG containing only the vertices  $\{t_1, y_1, t_2, y_2\}$  which represents the independence relations  $(\ddagger)$  and does not also imply some other independence relation that is not implied by  $\mathcal{D}_1$ . Consequently, any DAG model on these vertices will either fail to represent an independence relation, and hence contain ‘too many’ edges, or will impose some additional independence restriction that is not implied by  $\mathcal{D}_1$ .

Suppose that the patient’s underlying health status  $h$  is not observed, and the generating structure  $\mathcal{D}_1$  is unknown. In these circumstances, a conventional analysis would consider DAG models containing edges that are consistent with the known time order of the variables. Given sufficient data, any DAG imposing an extra independence relation will be rejected by a likelihood-ratio test, and a DAG representing some subset of the independence relations, such as the DAG in Figure 2(ii), will be chosen. However, any such graph will contain the extra edge  $t_1 \rightarrow y_2$ , and fail to represent the marginal independence of these variables. Thus such an analysis would conceal the fact that the first treatment does not affect the second outcome. This is also an undesirable result from a purely predictive perspective, since a model which incorporated this marginal independence constraint would be more parsimonious.

Moreover, even if we were to consider DAGs that were compatible with a non-temporal ordering of  $\{y_1, y_2, t_1, t_2\}$ , we would still be unable to find a DAG which represented all and only the independence relations in  $(\ddagger)$ . An analysis based on undirected graphs, or chain graphs, under the LWF global Markov property, would still include additional edges. (It is possible to represent the independence structure of  $\mathcal{D}_1$  via a chain graph with the AMP Markov property, but this does not hold for an arbitrary DAG under marginalization. See Section 9.4.)

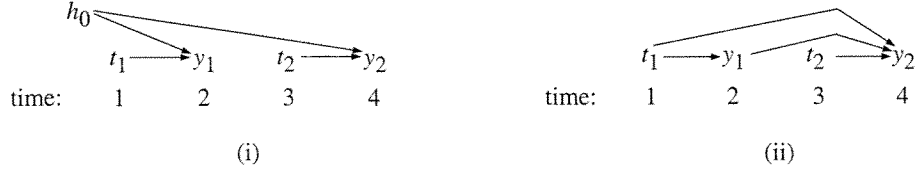


Figure 2: (i) a directed acyclic graph  $\mathcal{D}_1$ , representing a hypothesis concerning two completely randomized treatments and two outcomes (see text for further description); (ii) the DAG model  $\mathcal{D}_2$  resulting from a conventional analysis of  $\{t_1, y_1, t_2, y_2\}$ .

One response to this situation is to consider latent variable (LV) models, since  $h$  is a hidden variable in the model described by  $\mathcal{D}_1$ . Though this is certainly a possible approach in circumstances where much is known about the generating process, it seems unwise in other situations since LV models lack almost all of the desirable statistical properties attributed to graphical models (without hidden variables) above. In particular:

- LV models are not always identified;
- the likelihood may be multi-modal;
- any inference may be very sensitive to assumptions made about the unobserved variables;
- LV models with hidden variables have been proved *not* to be curved exponential families even in very simple cases (Geiger et al., 2001);
- LV models do not in general have a well-defined dimension for use in scores such as BIC, or  $\chi^2$ -tests (this follows from the previous point);
- the set of distributions associated with an LV model may be difficult to characterize (see Settini and Smith, 1999, 1998, Geiger et al., 2001, for recent results);
- LV models do not form a tractable search space: an arbitrary number of hidden variables may be incorporated, so the class contains infinitely many different structures relating a finite set of variables.

This presents the modeller with a dilemma: in many contexts it is clearly unrealistic to assume that there are no unmeasured confounding variables, and misleading analyses may result (as shown above). However, models that explicitly include hidden variables may be very hard to work with for the reasons just given.

The class of ancestral graph Markov models described in this paper is intended to provide a partial resolution to this conundrum. This class extends the class of DAG models, but is closed under marginalization. In addition, as we show in this paper, at least in the Gaussian case these models retain many of the desirable properties possessed by standard graphical models. It should be noted however that two different DAG models may lead to the same ancestral graph, so in this sense information is lost.

Up to this point we have considered closure under marginalization. There is a similar notion of closure under conditioning that is motivated by considering selection effects (see Cox and Wermuth, 1996, Cooper, 1995). UG Markov models are closed under conditioning, DAG models are not. The class of Markov models described here is also closed under conditioning.

The remainder of the paper is organized as follows:

We introduce basic graphical notation and definitions in Section 2. Section 3 introduces the class of ancestral graphs and the associated global Markov property. We also define the subclass of maximal ancestral graphs, which obey a pairwise Markov property.

In Section 4 we formally define the operation of marginalizing and conditioning for independence models, and a corresponding graphical transformation. Theorem 4.18 establishes that the independence model associated with the transformed graph is the same as the model resulting from applying the operations of marginalizing and conditioning to the independence model given by the original graph. It is also shown that the graphical transformations commute (Theorem 4.20).

Two extension results are proved in Section 5. It is first shown that by adding edges a non-maximal graph may be made maximal and this extension is unique (Theorem 5.1). Second, it is demonstrated that a maximal graph may be made complete (so that there is an edge between every pair of vertices) by a sequence of edge additions that preserve maximality (Theorem 5.6). In Section 6 it is shown that every maximal ancestral graph may be obtained by transforming a DAG, the structure of which bears a simple

relation to the original ancestral graph (Theorem 6.4). Consequently, every independence model associated with an ancestral graph may be obtained by applying the operations of marginalizing and conditioning to some independence model given by a DAG.

Section 7 relates the operations of marginalizing and conditioning that have been defined for independence models to probability distributions. Theorem 7.6 then shows that the global Markov property for ancestral graphs is complete.

In Section 8 we define a Gaussian parametrization of an ancestral graph. It is shown in Theorem 8.7 that each parameter is either a concentration, a regression coefficient, or a residual variance or covariance. Theorem 8.14 establishes that if the graph is maximal then the set of Gaussian distributions associated with the parametrization is exactly the set of Gaussian distributions which obey the global Markov property for the graph.

Section 9 contrasts the class of ancestral graphs to summary graphs, introduced by Wermuth et al. (1994), and MC-graphs introduced by Koster (1999a). Finally Section 10 contains a brief discussion.

## 2 Basic Definitions and Concepts

In this section we introduce notation and terminology for describing independence models and graphs.

### 2.1 Independence Models

An *independence model*  $\mathfrak{I}$  over a set  $V$  is a set of triples  $\langle X, Y \mid Z \rangle$  where  $X$ ,  $Y$  and  $Z$  are disjoint subsets of  $V$ ;  $X$  and  $Y$  are non-empty. The triple  $\langle X, Y \mid Z \rangle$  is interpreted as saying that  $X$  is independent of  $Y$  given  $Z$ . In Section 7 we relate this definition to conditional independence in a probability distribution. (As defined here, an ‘independence model’ need not correspond to the set of independence relations holding in any probability distribution.)

#### 2.1.1 Graphical independence models

A *graph*  $\mathcal{G}$  is an ordered pair  $(V, E)$  where  $V$  is a set of vertices and  $E$  is a set of edges. A *separation criterion*  $C$  associates an independence model  $\mathfrak{I}_C(\mathcal{G})$

with graph  $\mathcal{G}$ :

$$\langle X, Y \mid Z \rangle \in \mathfrak{I}_C(\mathcal{G}) \iff X \text{ is separated from } Y \text{ by } Z \text{ in } \mathcal{G} \text{ under criterion } C$$

Such a criterion  $C$  is also referred to as a *global Markov property*. The d-separation criterion introduced by Pearl (1988) is an example of such a criterion.

## 2.2 Mixed Graphs

A *mixed graph* is a graph containing three types of edge, *undirected*,  $(-)$ , *directed*,  $(\rightarrow)$ , and *bi-directed*  $(\leftrightarrow)$ . We use the following terminology to describe relations between variables in such a graph:

$$\text{If } \left\{ \begin{array}{l} \alpha - \beta \\ \alpha \leftrightarrow \beta \\ \alpha \rightarrow \beta \\ \alpha \leftarrow \beta \end{array} \right\} \text{ in } \mathcal{G} \text{ then } \alpha \text{ is a } \left\{ \begin{array}{l} \text{neighbour} \\ \text{spouse} \\ \text{parent} \\ \text{child} \end{array} \right\} \text{ of } \beta, \text{ and } \left\{ \begin{array}{l} \alpha \in \text{neg}(\beta) \\ \alpha \in \text{sp}_{\mathcal{G}}(\beta) \\ \alpha \in \text{pa}_{\mathcal{G}}(\beta) \\ \alpha \in \text{ch}_{\mathcal{G}}(\beta) \end{array} \right\}.$$

Note that the three edge types should be considered as distinct symbols, and in particular,

$$\alpha - \beta \neq \alpha \rightrightarrows \beta \neq \alpha \leftrightarrow \beta.$$

If there is an edge  $\alpha \rightarrow \beta$ , or  $\alpha \leftrightarrow \beta$  then there is said to be *an arrowhead at  $\beta$  on this edge*. If there is at least one edge between a pair of vertices then these vertices are *adjacent*. We do not allow a vertex to be adjacent to itself.

A graph  $\mathcal{G}' = (V', E')$  is a *subgraph* of  $\mathcal{G} = (V, E)$  if  $V' \subseteq V$  and every edge in  $\mathcal{G}'$  is present in  $\mathcal{G}$ . The *induced subgraph of  $\mathcal{G}$  over  $A$* , denoted  $\mathcal{G}_A$  has vertex set  $A$ , and contains every edge present in  $\mathcal{G}$  between the vertices in  $A$ . (See Appendix A.1 for more formal statements of these definitions.)

## 2.3 Paths and Edge Sequences

A *sequence of edges* between  $\alpha$  and  $\beta$  in  $\mathcal{G}$  is an ordered (multi)set of edges  $\langle \epsilon_1, \dots, \epsilon_n \rangle$ , such that there exists a sequence of vertices (not necessarily distinct)  $\langle \alpha \equiv \omega_1, \dots, \omega_{n+1} \equiv \beta \rangle (n \geq 0)$ , where edge  $\epsilon_i$  has endpoints  $\omega_i, \omega_{i+1}$ . A sequence of edges for which the corresponding sequence of vertices contains no repetitions is called a *path*. We will use bold greek ( $\boldsymbol{\mu}$ ) to denote paths and single edges, and fraktur ( $\mathfrak{s}$ ) to denote sequences. Note that the



result of concatenating two paths with a common endpoint is not necessarily a path, though it is always a sequence. Paths and sequences consisting of a single vertex, corresponding to a sequence of no edges, are permitted for the purpose of simplifying proofs; such paths will be called *empty* as the set of associated edges is empty.

We denote a *subpath* of a path  $\pi$ , by  $\pi(\omega_j, \omega_{k+1}) \equiv \langle \epsilon_j, \dots, \epsilon_k \rangle$ , and likewise for sequences. Unlike a subpath, a subsequence is not uniquely specified by the start and end vertices, hence the context will also make clear which occurrence of each vertex in the sequence is referred to.

We define a path as a sequence of edges rather than vertices because the latter does not specify a unique path when there may be two edges between a given pair of vertices. (However, from Section 3 on we will only consider graphs containing at most one edge between each pair of vertices.) A path of the form  $\alpha \rightarrow \dots \rightarrow \beta$ , on which every edge is of the form  $\rightarrow$ , with the arrowheads pointing towards  $\beta$  is a *directed path from  $\alpha$  to  $\beta$* .

## 2.4 Ancestors and Anterior Vertices

A vertex  $\alpha$  is said to be an *ancestor* of a vertex  $\beta$  if *either* there is a directed path  $\alpha \rightarrow \dots \rightarrow \beta$  from  $\alpha$  to  $\beta$ , *or*  $\alpha = \beta$ .

A vertex  $\alpha$  is said to be *anterior* to a vertex  $\beta$  if there is a path  $\mu$  on which every edge is either of the form  $\gamma - \delta$ , or  $\gamma \rightarrow \delta$  with  $\delta$  between  $\gamma$  and  $\beta$ , or  $\alpha = \beta$ ; i.e. there are no edges  $\gamma \leftrightarrow \delta$  and there are no edges  $\gamma \leftarrow \delta$  pointing towards  $\alpha$ . Such a path is said to be an *anterior path* from  $\alpha$  to  $\beta$ .

We apply these definitions disjunctively to sets:

$$\begin{aligned} an(X) &= \{\alpha \mid \alpha \text{ is an ancestor of } \beta \text{ for some } \beta \in X\}; \\ ant(X) &= \{\alpha \mid \alpha \text{ is anterior to } \beta \text{ for some } \beta \in X\}. \end{aligned}$$

Our usage of the terms ‘ancestor’ and ‘anterior’ differs from Lauritzen (1996), but follows Frydenberg (1990a).

**Proposition 2.1** *In a mixed graph  $\mathcal{G}$*

- (i) *If  $X \subseteq Y$  then  $an(X) \subseteq an(Y)$  and  $ant(X) \subseteq ant(Y)$ ;*
- (ii)  *$X \subseteq ant(X) = ant(ant(X))$  and  $X \subseteq an(X) = an(an(X))$ ;*
- (iii)  *$an(X \cup Y) = an(X) \cup an(Y)$  and  $ant(X \cup Y) = ant(X) \cup ant(Y)$ .*

**Proof:** These properties follow directly from the definitions of  $\text{an}(\cdot)$  and  $\text{ant}(\cdot)$ .  $\square$

**Proposition 2.2** *If  $X$  and  $Y$  are disjoint sets of vertices in a mixed graph  $\mathcal{G}$  then*

- (i)  $\text{ant}(\text{ant}(X) \setminus Y) = \text{ant}(X)$ ;
- (ii)  $\text{an}(\text{an}(X) \setminus Y) = \text{an}(X)$ .

**Proof:** (i) Since  $X$  and  $Y$  are disjoint,  $X \subseteq \text{ant}(X) \setminus Y$ . By Proposition 2.1(i),  $\text{ant}(X) \subseteq \text{ant}(\text{ant}(X) \setminus Y)$ . Conversely,  $\text{ant}(X) \setminus Y \subseteq \text{ant}(X)$  so  $\text{ant}(\text{ant}(X) \setminus Y) \subseteq \text{ant}(\text{ant}(X)) = \text{ant}(X)$ , by Proposition 2.1(i) and (ii).  $\square$

The proof of (ii) is very similar.  $\square$

A directed path from  $\alpha$  to  $\beta$  together with an edge  $\beta \rightarrow \alpha$  is called a *(fully) directed cycle*. An anterior path from  $\alpha$  to  $\beta$  together with an edge  $\beta \rightarrow \alpha$  is called a *partially directed cycle*. A *directed acyclic graph (DAG)* is a mixed graph in which all edges are directed, and there are no directed cycles.

### 3 Ancestral Graphs

The class of mixed graphs is much larger than required for our purposes, in particular, under natural separation criteria, it includes independence models that do not correspond to DAG models under marginalizing and conditioning. We now introduce the subclass of ancestral graphs.

#### 3.1 Definition of an ancestral graph

An *ancestral* graph  $\mathcal{G}$  is a mixed graph in which the following conditions hold for all vertices  $\alpha$  in  $\mathcal{G}$ :

- (i)  $\alpha \notin \text{ant}(\text{pa}(\alpha) \cup \text{sp}(\alpha))$ ;
- (ii) if  $\text{ne}(\alpha) \neq \emptyset$  then  $\text{pa}(\alpha) \cup \text{sp}(\alpha) = \emptyset$ .

In words, condition (i) requires that if  $\alpha$  and  $\beta$  are joined by an edge with an arrowhead at  $\alpha$ , then  $\alpha$  is not anterior to  $\beta$ . Condition (ii) requires that there be no arrowheads present at a vertex which is an endpoint of an undirected edge. Condition (i) implies that if  $\alpha$  and  $\beta$  are joined by an edge with an arrowhead at  $\alpha$ , then  $\alpha$  is not an ancestor of  $\beta$ . This is the motivation for terming such graphs ‘ancestral’. (See also Corollary 3.10.) Examples of ancestral and non-ancestral mixed graphs are shown in Figure 3.



Figure 3: (a) Mixed graphs that are not ancestral; (b) ancestral mixed graphs.

**Lemma 3.1** *In an ancestral graph for every vertex  $\alpha$  the sets  $\text{ne}(\alpha)$ ,  $\text{pa}(\alpha)$ ,  $\text{ch}(\alpha)$  and  $\text{sp}(\alpha)$  are disjoint, thus there is at most one edge between any pair of vertices.*

**Proof:**  $\text{ne}(\alpha)$ ,  $\text{pa}(\alpha)$  and  $\text{ch}(\alpha)$  are disjoint by condition (i).  $\text{ne}(\alpha) \cap \text{sp}(\alpha) = \emptyset$  by (ii) since at most one of these sets is non-empty. Finally (i) implies that  $\text{sp}(\alpha) \cap \text{pa}(\alpha) \subseteq \text{sp}(\alpha) \cap \text{ant}(\alpha) = \emptyset$ , and likewise  $\text{sp}(\alpha) \cap \text{ch}(\alpha) = \emptyset$ .  $\square$

**Lemma 3.2** *If  $\mathcal{G}$  is an ancestral graph then the following hold:*

- (a) *If  $\alpha$  and  $\beta$  are adjacent in  $\mathcal{G}$  and  $\alpha \in \text{an}(\beta)$  then  $\alpha \rightarrow \beta$ .*
- (b) *The configurations  $\alpha - \beta \leftrightarrow \gamma$  and  $\alpha - \beta \leftarrow \gamma$  do not occur (regardless of whether  $\alpha$  and  $\gamma$  are adjacent).*
- (c) *There are no directed cycles or partially directed cycles.*

**Proof:** (a) follows because condition (i) rules out  $\alpha \leftarrow \beta$  or  $\alpha \leftrightarrow \beta$ , while (ii) rules out  $\alpha - \beta$ . (b) is simply a restatement of condition (ii). (c) follows because (i) rules out fully directed cycles, while the configuration  $\rightarrow \gamma -$  occurs in any partially directed cycle.  $\square$

If there is at most one edge between two vertices in a graph then conditions (a), (b) and (c) in Lemma 3.2 are sufficient for  $\mathcal{G}$  to be ancestral.

**Corollary 3.3** *In an ancestral graph an anterior path from  $\alpha$  to  $\beta$  takes one of three forms:  $\alpha - \dots - \beta$ ,  $\alpha \rightarrow \dots \rightarrow \beta$ , or  $\alpha - \dots - \rightarrow \dots \rightarrow \beta$ .*

**Proof:** Follows from the definition of an anterior path and Lemma 3.2(b).  $\square$

**Proposition 3.4** *If  $\mathcal{G}$  is an undirected graph, or a directed acyclic graph, then  $\mathcal{G}$  is an ancestral graph.*

**Proposition 3.5** *If  $\mathcal{G}$  is an ancestral graph and  $\mathcal{G}'$  is a subgraph of  $\mathcal{G}$ , then  $\mathcal{G}'$  is ancestral.*

**Proof:** The definition of an ancestral graph only forbids certain configurations of edges. If these do not occur in  $\mathcal{G}$  then they do not occur in a subgraph  $\mathcal{G}'$ .  $\square$

## 3.2 Undirected Edges in an Ancestral Graph

Let  $\text{un}_{\mathcal{G}} \equiv \{\alpha \mid \text{pa}_{\mathcal{G}}(\alpha) \cup \text{sp}_{\mathcal{G}}(\alpha) = \emptyset\}$ , be the set of vertices at which no arrowheads are present in  $\mathcal{G}$ . Note that if  $\text{ne}_{\mathcal{G}}(\alpha) \neq \emptyset$  then, by condition (ii) in the definition of an ancestral graph,  $\alpha \in \text{un}_{\mathcal{G}}$ , so  $\text{un}_{\mathcal{G}}$  contains all endpoints of undirected edges in  $\mathcal{G}$ .

**Proposition 3.6** *If  $\mathcal{G}$  is an ancestral graph, and  $\mathcal{G}'$  is a subgraph with the same vertex set, then  $\text{un}_{\mathcal{G}} \subseteq \text{un}_{\mathcal{G}'}$ .*

**Proof:** Since  $\mathcal{G}'$  has a subset of the edges in  $\mathcal{G}$ ,  $\text{pa}_{\mathcal{G}}(\alpha) \cup \text{sp}_{\mathcal{G}}(\alpha) = \emptyset$  implies  $\text{pa}_{\mathcal{G}'}(\alpha) \cup \text{sp}_{\mathcal{G}'}(\alpha) = \emptyset$ .  $\square$

**Lemma 3.7** *If  $\mathcal{G}$  is an ancestral graph with vertex set  $V$*

$$\text{and } \left\{ \begin{array}{l} \alpha \leftrightarrow \beta \\ \alpha - \beta \\ \alpha \rightarrow \beta \end{array} \right\} \text{ in } \mathcal{G} \text{ then } \left\{ \begin{array}{l} \alpha, \beta \in V \setminus \text{un}_{\mathcal{G}} \\ \alpha, \beta \in \text{un}_{\mathcal{G}} \\ \beta \in V \setminus \text{un}_{\mathcal{G}} \end{array} \right\}.$$

**Proof:** Follows directly from definition of  $\text{un}_{\mathcal{G}}$  and Lemma 3.2(b).  $\square$

Lemma 3.7 shows that any ancestral graph can be split into an undirected graph  $\mathcal{G}_{\text{un}_{\mathcal{G}}}$ , and an ancestral graph containing no undirected edges  $\mathcal{G}_{V \setminus \text{un}_{\mathcal{G}}}$ ; any edge between a vertex  $\alpha \in \text{un}_{\mathcal{G}}$  and a vertex  $\beta \in V \setminus \text{un}_{\mathcal{G}}$  takes the form  $\alpha \rightarrow \beta$ . See Figure 4. This result is useful in developing parametrizations for the resulting independence models (see Section 8).

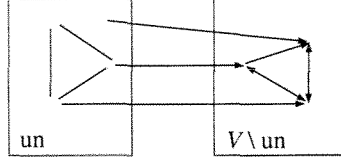


Figure 4: Schematic showing decomposition of an ancestral graph into an undirected graph and a graph containing no undirected edges.

**Lemma 3.8** *For an ancestral graph  $\mathcal{G}$ ,*

- (i) *if  $\alpha \in \text{un}_{\mathcal{G}}$  then  $\beta \in \text{ant}_{\mathcal{G}}(\alpha) \Rightarrow \alpha \in \text{ant}_{\mathcal{G}}(\beta)$ ;*
- (ii) *if  $\alpha$  and  $\beta$  are such that  $\alpha \neq \beta$ ,  $\alpha \in \text{ant}_{\mathcal{G}}(\beta)$  and  $\beta \in \text{ant}_{\mathcal{G}}(\alpha)$  then  $\alpha, \beta \in \text{un}_{\mathcal{G}}$ , and there is a path joining  $\alpha$  and  $\beta$  on which every edge is undirected;*
- (iii)  $\text{ant}_{\mathcal{G}}(\alpha) \setminus \text{an}_{\mathcal{G}}(\alpha) \subseteq \text{un}_{\mathcal{G}}$ .

**Proof:** (i) follows from Lemma 3.2(b) and Corollary 3.3. (ii) follows since by Lemma 3.2(c) there are no partially directed cycles and thus the anterior paths between  $\alpha$  and  $\beta$  consist only of undirected edges, so  $\alpha, \beta \in \text{un}_{\mathcal{G}}$  by Lemma 3.7. (iii) follows because if a vertex  $\beta$  is anterior to  $\alpha$ , but not an ancestor of  $\alpha$ , then by Corollary 3.3 any anterior path starts with an undirected edge, and the result follows from Lemma 3.7.  $\square$

**Lemma 3.9** *If  $\mathcal{G}$  is an ancestral graph, and  $\alpha, \beta$  are adjacent vertices in  $\mathcal{G}$  then:*

- (i)  $\alpha - \beta \Leftrightarrow \alpha \in \text{ant}_{\mathcal{G}}(\beta), \beta \in \text{ant}_{\mathcal{G}}(\alpha)$ ;
- (ii)  $\alpha \rightarrow \beta \Leftrightarrow \alpha \in \text{ant}_{\mathcal{G}}(\beta), \beta \notin \text{ant}_{\mathcal{G}}(\alpha)$ ;
- (iii)  $\alpha \leftrightarrow \beta \Leftrightarrow \alpha \notin \text{ant}_{\mathcal{G}}(\beta), \beta \notin \text{ant}_{\mathcal{G}}(\alpha)$ .

**Proof:** (i) ( $\Rightarrow$ ) follows by the definition of anterior; (i) ( $\Leftarrow$ ) by Lemma 3.8 (ii) and Lemma 3.2(b). Claim (ii)( $\Rightarrow$ ) follows by the definition of anterior and property (i) of an ancestral graph; (ii)( $\Leftarrow$ ) follows because from Lemma 3.8 (i),  $\beta \notin \text{un}_{\mathcal{G}}$ , and so by Lemma 3.7 and property (i) of an ancestral graph,  $\alpha \rightarrow \beta$ . (iii)( $\Rightarrow$ ) follows by property (i) of an ancestral graph. (iii)( $\Leftarrow$ ) follows by definition of anterior.  $\square$

A direct consequence of Lemma 3.9 is that an ancestral graph is uniquely determined by its adjacencies (or ‘skeleton’) and anterior relations. More formally:

**Corollary 3.10** *If  $\mathcal{G}_1$  and  $\mathcal{G}_2$  are two ancestral graphs with the same vertex set  $V$ , and adjacencies, then if  $\forall \alpha, \beta \in V$ , adjacent in  $\mathcal{G}_1$  and  $\mathcal{G}_2$ ,*

$$\alpha \in \text{ant}_{\mathcal{G}_1}(\beta) \Leftrightarrow \alpha \in \text{ant}_{\mathcal{G}_2}(\beta)$$

*then  $\mathcal{G}_1 = \mathcal{G}_2$ .*

**Proof:** Follows directly from Lemma 3.9. □

Note that this does not hold in general for non-ancestral graphs. See Figure 5 for an example.

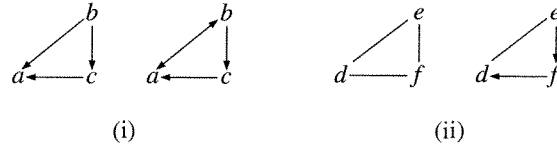


Figure 5: Two pairs of graphs that share the same adjacencies and anterior relations between adjacent vertices, and yet are not equivalent.

### 3.3 Bi-Directed Edges in an Ancestral Graph

The following Lemma shows that the ancestor relation induces a partial ordering on the bi-directed edges in an ancestral graph.

**Lemma 3.11** *Let  $\mathcal{G}$  be an ancestral graph. The relation  $\prec$  defined by:*

$$\alpha \leftrightarrow \beta \prec \gamma \leftrightarrow \delta \text{ if } \alpha, \beta \in \text{an}(\{\gamma, \delta\}) \text{ and } \{\alpha, \beta\} \neq \{\gamma, \delta\}$$

*defines a strict (irreflexive) partial order on the bi-directed edges in  $\mathcal{G}$ .*

**Proof:** Transitivity of the relation  $\prec$  follows directly from transitivity of the ancestor relation. Suppose for a contradiction that  $\alpha \leftrightarrow \beta \prec \gamma \leftrightarrow \delta \prec \alpha \leftrightarrow \beta$ , but  $\{\alpha, \beta\} \neq \{\gamma, \delta\}$ . Either  $\alpha \notin \{\gamma, \delta\}$  or  $\beta \notin \{\gamma, \delta\}$ . Without loss of

generality, suppose the former. Since  $\alpha \in \text{an}(\{\gamma, \delta\})$  and  $\gamma, \delta \in \text{an}(\{\alpha, \beta\})$  it then follows that either  $\alpha \in \text{an}(\beta)$ , or there is a directed cycle containing  $\alpha$  and  $\gamma$  or  $\delta$ . In both cases condition (i) in the definition of an ancestral graph is violated.  $\square$

Note that the relation given by

$$\alpha \leftrightarrow \beta \prec^* \gamma \leftrightarrow \delta \text{ if } \left( \alpha \in \text{an}(\{\gamma, \delta\}) \text{ or } \beta \in \text{an}(\{\gamma, \delta\}) \right) \text{ and } \{\alpha, \beta\} \neq \{\gamma, \delta\}$$

does *not* give an ordering on the bi-directed edges as shown by the ancestral graph in Figure 6. This is significant since it means that in an ancestral graph it is not possible in general to construct ordered blocks of vertices such that all bi-directed edges are within blocks and all directed edges are between vertices in different blocks and are directed in accordance with the ordering.



Figure 6: An ancestral graph which cannot be arranged in ordered blocks with bi-directed edges within blocks and edges between blocks directed in accordance with the ordering. (See text for further discussion.)

### 3.4 The Pathwise m-separation Criterion

We now extend Pearl's d-separation criterion (see Pearl, 1988), defined originally for DAGs, to ancestral graphs.

A non-endpoint vertex  $\zeta$  on a path is a *collider on the path* if the edges preceding and succeeding  $\zeta$  on the path have an arrowhead at  $\zeta$ , i.e.  $\rightarrow \zeta \leftarrow$ ,  $\leftrightarrow \zeta \leftrightarrow$ ,  $\leftrightarrow \zeta \leftarrow$ ,  $\rightarrow \zeta \leftrightarrow$ . A non-endpoint vertex  $\zeta$  on a path which is not a collider is a *non-collider on the path*. A path between vertices  $\alpha$  and  $\beta$  in an ancestral graph  $\mathcal{G}$  is said to be *m-connecting given a set  $Z$*  (possibly empty), with  $\alpha, \beta \notin Z$ , if

- (i) every non-collider on the path is not in  $Z$ , and
- (ii) every collider on the path is in  $\text{ant}_{\mathcal{G}}(Z)$ .

If there is no path  $m$ -connecting  $\alpha$  and  $\beta$  given  $Z$ , then  $\alpha$  and  $\beta$  are said to be  $m$ -separated given  $Z$ . Sets  $X$  and  $Y$  are  $m$ -separated given  $Z$ , if for every pair  $\alpha, \beta$ , with  $\alpha \in X$  and  $\beta \in Y$ ,  $\alpha$  and  $\beta$  are  $m$ -separated given  $Z$  ( $X, Y, Z$  are disjoint sets;  $X, Y$  are non-empty). We denote the independence model resulting from applying the  $m$ -separation criterion to  $\mathcal{G}$ , by  $\mathfrak{I}_m(\mathcal{G})$ .

This is an extension of Pearl's  $d$ -separation criterion to mixed graphs in that in a DAG  $\mathcal{D}$ , a path is  $d$ -connecting if and only if it is  $m$ -connecting. See Figure 7(a) for an example. The formulation of this property leads directly to:

**Proposition 3.12** *If  $\mathcal{G}$  is an ancestral graph, and  $\mathcal{G}'$  is a subgraph with the same vertex set, then  $\mathfrak{I}_m(\mathcal{G}) \subseteq \mathfrak{I}_m(\mathcal{G}')$ .*

**Proof:** This holds because any path in  $\mathcal{G}'$  exists in  $\mathcal{G}$ .  $\square$

Notice that it follows directly from Corollary 3.3, and Lemma 3.2 (b) that if  $\gamma$  is a collider on a path  $\pi$  in an ancestral graph  $\mathcal{G}$  then  $\gamma \in \text{ant}_{\mathcal{G}}(\beta) \Leftrightarrow \gamma \in \text{an}_{\mathcal{G}}(\beta)$ . Since the set of  $m$ -connecting paths will not change, strengthening condition (ii) in the definition of  $m$ -separation to:

(ii)' every collider on the path is in  $\text{an}_{\mathcal{G}}(Z)$ .

will not change the resulting independence model  $\mathfrak{I}_m(\mathcal{G})$ . This formulation is closer to the original definition of  $d$ -separation as originally defined for directed acyclic graphs, since it does not use the anterior relation. The only change is that the definitions of 'collider' and 'non-collider' have been extended to allow for edges of the form  $-$  and  $\leftrightarrow$ . (Also see the definition of 'h-separation' introduced in Verma and Pearl (1990).)

### 3.4.1 Properties of $m$ -connecting paths

We now prove two Lemmas giving properties of  $m$ -connecting paths that we will exploit in Section 3.6.

**Lemma 3.13** *If  $\pi$  is a path  $m$ -connecting  $\alpha$  and  $\beta$  given  $Z$  in an ancestral graph  $\mathcal{G}$  then every vertex on  $\pi$  is in  $\text{ant}(\{\alpha, \beta\} \cup Z)$ .*



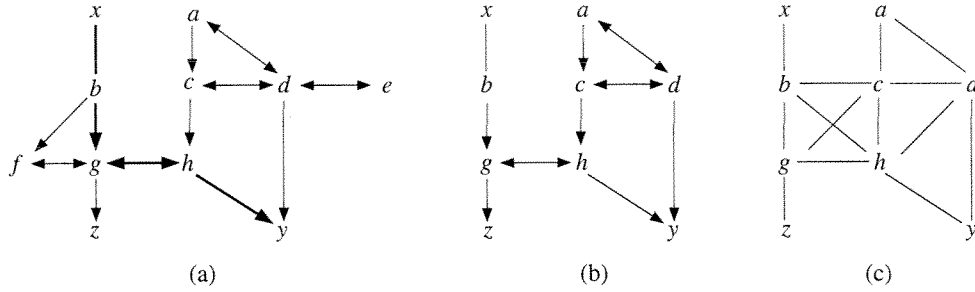


Figure 7: Example of global Markov properties. (a) An ancestral graph  $\mathcal{G}$ , thicker edges form a path  $m$ -connecting  $x$  and  $y$  given  $\{z\}$ ; (b) the subgraph  $\mathcal{G}_{\text{ant}(\{x,y,z\})}$ ; (c) the augmented graph  $(\mathcal{G}_{\text{ant}(\{x,y,z\})})^a$ , in which  $x$  and  $y$  are not separated by  $\{z\}$ .

**Proof:** Suppose  $\gamma$  is on  $\pi$  and is not anterior to  $\alpha$  or  $\beta$ . Then, on each of the subpaths  $\pi(\alpha, \gamma)$  and  $\pi(\gamma, \beta)$ , there is at least one edge with an arrowhead pointing towards  $\gamma$  along the subpath. Let  $\phi_{\alpha\gamma}$  and  $\phi_{\gamma\beta}$  be the vertices at which such arrowheads occur that are closest to  $\gamma$  on the respective subpaths. There are now three cases:

If  $\gamma \neq \phi_{\gamma\beta}$  then  $\pi(\gamma, \phi_{\gamma\beta})$  is an anterior path from  $\gamma$  to  $\phi_{\gamma\beta}$ . It further follows from Lemma 3.2(b) and Corollary 3.3 that  $\phi_{\gamma\beta}$  is a collider on  $\pi$ , hence anterior to  $Z$ , since  $\pi$  is  $m$ -connecting given  $Z$ . Hence  $\gamma \in \text{ant}(Z)$ .

If  $\gamma \neq \phi_{\alpha\gamma}$  then by a symmetric argument to the previous case it follows that  $\gamma$  is anterior to  $\phi_{\alpha\gamma}$ , and  $\phi_{\alpha\gamma}$  is a collider on  $\pi$  and thus anterior to  $Z$ . Thus in this case,  $\gamma \in \text{ant}(Z)$ .

If  $\phi_{\alpha\gamma} = \beta = \phi_{\gamma\beta}$  then  $\gamma$  is a collider on  $\pi$ , hence anterior to  $Z$ .

□

**Lemma 3.14** *Let  $\mathcal{G}$  be an ancestral graph containing disjoint sets of vertices  $X, Y, Z$  ( $Z$  may be empty). If there are vertices  $\alpha \in X$  and  $\beta \in Y$  joined by a path  $\mu$  on which no non-collider is in  $Z$  and every collider is in  $\text{ant}(X \cup Y \cup Z)$  then there exist vertices  $\alpha^* \in X, \beta^* \in Y$  such that  $\alpha^*$  and  $\beta^*$  are  $m$ -connected given  $Z$  in  $\mathcal{G}$ .*

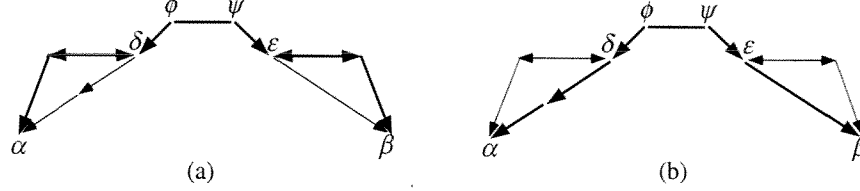


Figure 8: Illustration of Lemma 3.14: (a) a path on which every vertex is an ancestor of  $\alpha$  or  $\beta$ ; (b) a path m-connecting  $\alpha$  and  $\beta$  given  $\emptyset$ .

**Proof:** Let  $\mu^*$  be a path which contains the minimum number of colliders of any path between some vertex  $\alpha^* \in X$  and some vertex  $\beta^* \in Y$  on which no non-collider is in  $Z$  and every collider is in  $\text{ant}(X \cup Y \cup Z)$ .  $\mu^*$  is guaranteed to exist since the path  $\mu$  described in the Lemma has this form. In order to show that  $\mu^*$  m-connects  $\alpha^*$  and  $\beta^*$  given  $Z$  it is sufficient to show that every collider on  $\mu^*$  is in  $\text{ant}(Z)$ .

Suppose for a contradiction that there is a collider  $\gamma$  on  $\mu^*$  and  $\gamma \notin \text{ant}(Z)$ . By construction  $\gamma \in \text{ant}(X \cup Y \cup Z)$ , so either  $\gamma \in \text{ant}(X) \setminus \text{ant}(Z)$  or  $\gamma \in \text{ant}(Y) \setminus \text{ant}(Z)$ . Suppose the former, then there is a directed path  $\pi$  from  $\gamma$  to some vertex  $\alpha' \in X$ . Let  $\delta$  be the vertex closest to  $\beta^*$  on  $\mu^*$  which is also on  $\pi$ . By construction the paths  $\mu^*(\delta, \beta^*)$  and  $\pi(\delta, \alpha')$  do not intersect except at  $\delta$ . Hence concatenating these subpaths forms a path which satisfies the conditions on  $\mu^*$  but has fewer colliders than  $\mu^*$ , which is a contradiction. The case where  $\gamma \in \text{ant}(Y) \setminus \text{ant}(Z)$  is symmetric.  $\square$

**Corollary 3.15** *In an ancestral graph  $\mathcal{G}$ , there is a path  $\mu$  between  $\alpha$  and  $\beta$  on which no non-collider is in a set  $Z$  ( $\alpha, \beta \notin Z$ ) and every collider is in  $\text{ant}(\{\alpha, \beta\} \cup Z)$  if and only if there is a path m-connecting  $\alpha$  and  $\beta$  given  $Z$  in  $\mathcal{G}$ .*

**Proof:** One direction is immediate and the other is a special case of Lemma 3.14 with  $X = \{\alpha\}$ ,  $Y = \{\beta\}$ .  $\square$

This Corollary shows that condition (ii) in the definition of m-separation can be weakened to:

(ii)" every collider on the path is in  $\text{ant}(\{\alpha, \beta\} \cup Z)$ .

without changing the resulting independence model (for ancestral graphs).

### 3.4.2 Formulation via sequences

Koster (2000) shows that if the separation criterion is applied to sequences of edges (which may include repetitions of the same edge) as opposed to paths, then some simplification is possible. Under this formulation vertices  $\alpha$  and  $\beta$  in a mixed graph  $\mathcal{G}$  are said to be  $m$ -connecting given a set  $Z$  if there is a sequence  $\mathfrak{s}$  for which

- (i)\* every non-collider on  $\mathfrak{s}$  is not in  $Z$ , and
- (ii)\* every collider on  $\mathfrak{s}$  is in  $Z$ .

The definitions of collider and non-collider remain unchanged, but are applied to edges occurring in sequences, so  $\alpha \rightarrow \beta \leftarrow \alpha$  forms a collider. Koster (2000) proves that this criterion is identical to the  $m$ -separation criterion defined here for paths: the proof is based on the fact that there is a directed path from a collider  $\gamma$  to a vertex  $\zeta \in Z$  if and only if there is a sequence of the form  $\gamma \rightarrow \cdots \rightarrow \zeta \leftarrow \cdots \leftarrow \gamma$ .

We do not make use of this criterion in this paper, as paths, rather than sequences, are fundamental to our main construction (see Section 4.2.3).

## 3.5 The Augmentation $m^*$ -separation Criterion

The global Markov property for DAGs may be formulated via separation in an undirected graph, obtained from the original DAG by first forming a subgraph and then adding undirected edges between non-adjacent vertices that share a common child, a process known as ‘moralizing’. (See Lauritzen (1996), p.47 for details.) In this subsection we formulate the global Markov property for ancestral mixed graphs in this way. In the next subsection the resulting independence model is shown to be equivalent to that obtained via  $m$ -separation. It is useful to have two formulations of the Markov property because some proofs are simpler using one while other proofs are simpler using the other.

### 3.5.1 The augmented graph $(\mathcal{G})^a$

Two vertices  $\alpha$  and  $\beta$  in an ancestral graph  $\mathcal{G}$  are said to be *collider connected* if there is a path from  $\alpha$  to  $\beta$  in  $\mathcal{G}$  on which every vertex except the endpoints is a collider, such a path is called a *collider path*. (Koster (1999b) refers to

such a path as a ‘pure collision path’.) Note that if there is a single edge between  $\alpha$  and  $\beta$  in the graph then  $\alpha$  and  $\beta$  are (vacuously) collider connected.

The *augmented graph*, denoted  $(\mathcal{G})^a$ , derived from the mixed graph  $\mathcal{G}$  is an undirected graph with the same vertex set as  $\mathcal{G}$  such that

$$\gamma - \delta \text{ in } (\mathcal{G})^a \Leftrightarrow \gamma \text{ and } \delta \text{ are collider connected in } \mathcal{G}.$$

### 3.5.2 Definition of $m^*$ -separation

Sets  $X, Y$  and  $Z$  are said to be  $m^*$ -separated if  $X$  and  $Y$  are separated by  $Z$  in  $(\mathcal{G}_{\text{ant}(X \cup Y \cup Z)})^a$  ( $X, Y, Z$  are disjoint sets;  $X, Y$  are non-empty). Otherwise  $X$  and  $Y$  are said to be  $m^*$ -connected given  $Z$ . The resulting independence model is denoted by  $\mathfrak{I}_{m^*}(\mathcal{G})$ . See Figure 7(b),(c) for an example.

When applied to DAGs, or UGs, the augmentation criterion presented here is equivalent to the Lauritzen-Wermuth-Frydenberg moralization criterion. (See Section 9.4 for discussion of chain graphs.)

### 3.5.3 Minimal $m^*$ -connecting paths

If there is an edge  $\gamma - \delta$  in  $(\mathcal{G})^a$ , but there is no edge between  $\gamma$  and  $\delta$  in  $\mathcal{G}$ , then the edge is said to be *augmented*. A path connecting  $x$  and  $y$  given  $Z$  is said to be *minimal* if there is no other such path which connects  $x$  and  $y$  given  $Z$  but has fewer edges than  $\mu$ .

We now prove a property of minimal paths that is used in the next section:

**Lemma 3.16** *Let  $\mathcal{G}$  be an ancestral graph. If  $\mu$  is a minimal path connecting  $\alpha$  and  $\beta$  given  $Z$  in  $(\mathcal{G})^a$ , then a collider path in  $\mathcal{G}$  associated with an augmented edge  $\gamma - \delta$  on  $\mu$  has no vertex in common with  $\mu$ , or any collider path associated with another augmented edge on  $\mu$ , except possibly  $\gamma$  or  $\delta$ .*

**Proof:** Suppose that  $\gamma - \delta$  and  $\epsilon - \phi$  are two augmented edges, occurring in that order on  $\mu$ , and that the associated collider paths have in common a vertex which is not an endpoint of these paths. Then  $\gamma$  and  $\phi$  are adjacent in  $(\mathcal{G})^a$ . Thus a shorter path may be constructed by concatenating  $\mu(\alpha, \gamma)$ ,  $\gamma - \phi$  and  $\mu(\phi, \beta)$ , which is a contradiction. Likewise suppose that  $\kappa$  is a vertex on a collider path between  $\gamma$  and  $\delta$  which also occurs on  $\mu$ .  $\kappa$  either occurs before or after  $\gamma$  on the path. Suppose the former, then since  $\kappa - \delta$  in  $(\mathcal{G})^a$ , a shorter path may be formed by concatenating  $\mu(\alpha, \kappa)$ ,  $\kappa - \delta$  and  $\mu(\delta, \beta)$ . The case where  $\kappa$  occurs after  $\delta$  is similar.  $\square$

### 3.6 Equivalence of m-separation and m\*-separation

**Lemma 3.17** *In an ancestral graph  $\mathcal{G}$  suppose that  $\mu$  is a path which m-connects  $\alpha$  and  $\beta$  given  $Z$ . The sequence of non-colliders on  $\mu$  forms a path connecting  $\alpha$  and  $\beta$  in  $(\mathcal{G}_{\text{ant}(\{\alpha, \beta\} \cup Z)})^a$ .*

**Proof:** By Lemma 3.13 all the vertices on  $\mu$  are in  $\mathcal{G}_{\text{ant}(\{\alpha, \beta\} \cup Z)}$ . Suppose that  $\omega_i$  and  $\omega_{i+1}$  ( $1 \leq i \leq k-1$ ) are the successive non-colliders on  $\mu$ . The subpath  $\mu(\omega_i, \omega_{i+1})$  consists entirely of colliders, hence  $\omega_i$  and  $\omega_{i+1}$  are adjacent in  $(\mathcal{G}_{\text{ant}(\{\alpha, \beta\} \cup Z)})^a$ . Similarly  $\omega_1$  and  $\omega_k$  are adjacent to  $\alpha$  and  $\beta$  respectively in  $(\mathcal{G}_{\text{ant}(\{\alpha, \beta\} \cup Z)})^a$ .  $\square$

**Theorem 3.18** *For an ancestral graph  $\mathcal{G}$ ,  $\mathcal{I}_{m^*}(\mathcal{G}) = \mathcal{I}_m(\mathcal{G})$ .*

We break the proof into two pieces:

**Proof:**  $\mathcal{I}_{m^*}(\mathcal{G}) \subseteq \mathcal{I}_m(\mathcal{G})$

We proceed by showing that if  $\langle X, Y \mid Z \rangle \notin \mathcal{I}_m(\mathcal{G})$  then  $\langle X, Y \mid Z \rangle \notin \mathcal{I}_{m^*}(\mathcal{G})$ . If  $\langle X, Y \mid Z \rangle \notin \mathcal{I}_m(\mathcal{G})$  then there are vertices  $\alpha \in X$ ,  $\beta \in Y$  such that there is an m-connecting path  $\mu$  between  $\alpha$  and  $\beta$  given  $Z$  in  $\mathcal{G}$ . By Lemma 3.17 the non-colliders on  $\mu$  form a path  $\mu^*$  connecting  $\alpha$  and  $\beta$  in  $(\mathcal{G}_{\text{ant}(X \cup Y \cup Z)})^a$ . Since  $\mu$  is m-connecting, no non-collider on  $\mu$  is in  $Z$  hence no vertex on  $\mu^*$  is in  $Z$ . Thus  $\langle X, Y \mid Z \rangle \notin \mathcal{I}_{m^*}(\mathcal{G})$ .  $\square$

**Proof:**  $\mathcal{I}_m(\mathcal{G}) \subseteq \mathcal{I}_{m^*}(\mathcal{G})$

We show that if  $\langle X, Y \mid Z \rangle \notin \mathcal{I}_{m^*}(\mathcal{G})$  then  $\langle X, Y \mid Z \rangle \notin \mathcal{I}_m(\mathcal{G})$ . If  $\langle X, Y \mid Z \rangle \notin \mathcal{I}_{m^*}(\mathcal{G})$  then there are vertices  $\alpha \in X$ ,  $\beta \in Y$  such that there is a minimal path  $\pi$  connecting  $\alpha$  and  $\beta$  in  $(\mathcal{G}_{\text{ant}(X \cup Y \cup Z)})^a$  on which no vertex is in  $Z$ . Our strategy is to replace each augmented edge on  $\pi$  with a corresponding collider path in  $\mathcal{G}_{\text{ant}(X \cup Y \cup Z)}$  and replace the other edges on  $\pi$  with the corresponding edge in  $\mathcal{G}$ . It follows from Lemma 3.16 that the resulting sequence of edges forms a path from  $\alpha$  to  $\beta$  in  $\mathcal{G}$ , which we denote  $\nu$ . Further, any non-collider on  $\nu$  is a vertex on  $\pi$  and hence not in  $Z$ . Finally, since all vertices in  $\nu$  are in  $\mathcal{G}_{\text{ant}(X \cup Y \cup Z)}$  it follows that every collider is in  $\text{ant}(X \cup Y \cup Z)$ . Thus by Lemma 3.14 there are vertices  $\alpha^* \in X$  and  $\beta^* \in Y$  such that  $\alpha^*$  and  $\beta^*$  are m-connected given  $Z$  in  $\mathcal{G}$ . Thus  $\langle X, Y \mid Z \rangle \notin \mathcal{I}_m(\mathcal{G})$ .  $\square$

### 3.7 Maximal Ancestral Graphs

Independence models described by DAGs and undirected graphs satisfy pairwise Markov properties with respect to these graphs, hence every missing edge corresponds to a conditional independence (see Lauritzen (1996), p.32). This is not true in general for an arbitrary ancestral graph, as shown by the graph in Figure 9 (a).

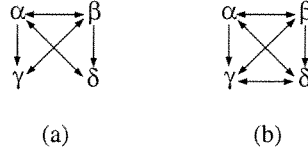


Figure 9: (a) The simplest example of a non-maximal ancestral graph:  $\gamma$  and  $\delta$  are not adjacent, but are m-connected given every subset of  $\{\alpha, \beta\}$ , hence  $\mathfrak{I}_m(\mathcal{G}) = \emptyset$ ; (b) an extension of the graph in (a) with the same (trivial) independence model.

This motivates the following definition: an ancestral graph  $\mathcal{G}$  is said to be *maximal* if for every pair of vertices  $\alpha, \beta$  if  $\alpha$  and  $\beta$  are not adjacent in  $\mathcal{G}$  then there is a set  $Z$  ( $\alpha, \beta \notin Z$ ), such that  $\langle \{\alpha\}, \{\beta\} \mid Z \rangle \in \mathfrak{I}_m(\mathcal{G})$ . Thus a graph is maximal if every missing edge corresponds to at least one independence in the corresponding independence model.

**Proposition 3.19** *If  $\mathcal{G}$  is an undirected graph, or a directed acyclic graph then  $\mathcal{G}$  is maximal.*

**Proof:** Follows directly from the existence of pairwise Markov properties for DAGs and undirected graphs.  $\square$

The use of the term ‘maximal’ is motivated by the following:

**Proposition 3.20** *If  $\mathcal{G} = (V, E)$  is a maximal ancestral graph, and  $\mathcal{G}$  is a subgraph of  $\mathcal{G}^* = (V, E^*)$ , then  $\mathfrak{I}_m(\mathcal{G}) = \mathfrak{I}_m(\mathcal{G}^*)$  implies  $\mathcal{G} = \mathcal{G}^*$ .*

**Proof:** If some pair  $\alpha, \beta$  are adjacent in  $\mathcal{G}^*$  but not  $\mathcal{G}$ , then in  $\mathcal{G}^*$ ,  $\alpha$  and  $\beta$  are m-connected by any subset of  $V \setminus \{\alpha, \beta\}$ . Hence  $\mathfrak{I}_m(\mathcal{G}) \neq \mathfrak{I}_m(\mathcal{G}^*)$ .  $\square$

Hence maximal ancestral graphs are maximal in the sense that no additional edge may be added to the graph without changing the independence model. The following Theorem gives the converse.

**Theorem 5.1** *If  $\mathcal{G}$  is an ancestral graph then there exists a unique maximal ancestral graph  $\bar{\mathcal{G}}$  formed by adding  $\leftrightarrow$  edges to  $\mathcal{G}$  such that  $\mathfrak{I}_m(\mathcal{G}) = \mathfrak{I}_m(\bar{\mathcal{G}})$ .*

We postpone the proof of this Theorem until Section 5.1, since it follows directly from another result. In Corollary 5.3 we show that a maximal ancestral graph satisfies the following:

**Pairwise Markov property**

*If there is no edge between  $\alpha$  and  $\beta$  in  $\mathcal{G}$  then*

$$\left\langle \{\alpha\}, \{\beta\} \mid \text{ant}(\{\alpha, \beta\}) \setminus \{\alpha, \beta\} \right\rangle \in \mathfrak{I}_m(\mathcal{G}).$$

### 3.8 Complete Ancestral Graphs

An ancestral graph is *complete* if there is an edge between every pair of distinct vertices. A graph is said to be *transitive* if  $\alpha \rightarrow \beta \rightarrow \gamma$  implies  $\alpha \rightarrow \gamma$ . Andersson et al. (1995, 1997), and Andersson and Perlman (1998) study properties of independence models based on transitive DAGs.

**Lemma 3.21** *If  $\mathcal{G}$  is a complete ancestral graph then*

- (i)  $\mathcal{G}$  is transitive;
- (ii) the induced subgraph  $\mathcal{G}_{\text{un}_\mathcal{G}}$  is a complete undirected graph;
- (iii) if  $\alpha \in V \setminus \text{un}_\mathcal{G}$  then  $\text{ant}_\mathcal{G}(\alpha) = \text{pa}_\mathcal{G}(\alpha) \cup \{\alpha\}$ ;
- (iv) if  $\alpha \in \text{un}_\mathcal{G}$  then  $\text{ant}_\mathcal{G}(\alpha) = \text{un}_\mathcal{G}$ .

**Proof:** If  $\alpha \rightarrow \beta \rightarrow \gamma$  in  $\mathcal{G}$  then  $\alpha \rightarrow \gamma$  since if  $\alpha - \gamma$ ,  $\alpha \leftarrow \gamma$ , or  $\alpha \leftrightarrow \gamma$  then  $\mathcal{G}$  would not be ancestral, establishing (i). If  $\alpha, \beta \in \text{un}_\mathcal{G}$  then by Lemma 3.7,  $\alpha - \beta$ , which establishes (ii). Suppose  $\alpha \in V \setminus \text{un}_\mathcal{G}$ ,  $\beta \in \text{ant}_\mathcal{G}(\alpha)$ . If  $\beta \in \text{un}_\mathcal{G}$  then  $\beta \rightarrow \alpha$ , by Lemma 3.7; if  $\beta \in V \setminus \text{un}_\mathcal{G}$  then  $\beta \in \text{an}_\mathcal{G}(\alpha)$  and so  $\beta \rightarrow \alpha$  by (i). Hence (iii) holds. (iv) follows directly from (ii).  $\square$

## 4 Marginalizing and Conditioning

In this section we first introduce marginalizing and conditioning for an independence model. We then define a graphical transformation of an ancestral graph. We show that the independence model corresponding to the transformed graph is the independence model obtained by marginalizing and conditioning the independence model of the original graph. In the remaining subsections we derive several useful consequences.

### 4.1 Marginalizing and Conditioning Independence Models ( $\mathfrak{I}_L^S$ )

An independence model  $\mathfrak{I}$  with vertex set  $V$  after marginalizing out a subset  $L$ , is simply the subset of triples which do not involve any vertices in  $L$ . More formally we define:

$$\mathfrak{I}_L \equiv \left\{ \langle X, Y \mid Z \rangle \mid \langle X, Y \mid Z \rangle \in \mathfrak{I}; (X \cup Y \cup Z) \cap L = \emptyset \right\}.$$

If  $\mathfrak{I}$  contains the independence relations present in a distribution  $P$ , then  $\mathfrak{I}_L$  contains the subset of independence relations remaining after marginalizing out the ‘Latent’ variables in  $L$ ; see Theorem 7.1. (Note the distinct uses of the vertical bar in  $\langle \cdot, \cdot \mid \cdot \rangle$  and  $\{ \cdot \mid \cdot \}$ .)

An independence model  $\mathfrak{I}$  with vertex set  $V$  after conditioning on a subset  $S$  is the set of triples defined as follows:

$$\mathfrak{I}^S \equiv \left\{ \langle X, Y \mid Z \rangle \mid \langle X, Y \mid Z \cup S \rangle \in \mathfrak{I}; (X \cup Y \cup Z) \cap S = \emptyset \right\}.$$

Thus if  $\mathfrak{I}$  contains the independence relations present in a distribution  $P$  then  $\mathfrak{I}^S$  constitutes the subset of independencies holding among the remaining variables after conditioning on  $S$ ; see Theorem 7.1. (Note that the set  $S$  is suppressed in the conditioning set in the independence relations in the resulting independence model.) The letter  $S$  is used because Selection effects represent one context in which conditioning may occur.

Combining these definitions we obtain:

$$\mathfrak{I}_L^S \equiv \left\{ \langle X, Y \mid Z \rangle \mid \langle X, Y \mid Z \cup S \rangle \in \mathfrak{I}; (X \cup Y \cup Z) \cap (S \cup L) = \emptyset \right\}.$$

**Proposition 4.1** *For an independence model  $\mathfrak{I}$  over  $V$  containing disjoint subsets  $S_1, S_2, L_1, L_2$ ,*



- (i)  $\mathfrak{I}_{\emptyset}^{\emptyset} = \mathfrak{I}$ ,
- (ii)  $\left(\mathfrak{I}_{L_1}^{S_1}\right)_{L_2}^{S_2} = \mathfrak{I}_{L_1 \cup L_2}^{S_1 \cup S_2}$ .

#### 4.1.1 Example

Consider the following independence model:

$$\mathfrak{I}^* = \left\{ \langle \{a, x\}, \{b, y\} \mid \{t\} \rangle, \langle \{a, x\}, \{b\} \mid \emptyset \rangle, \langle \{b, y\}, \{a\} \mid \emptyset \rangle, \langle \{a, b\}, \{t\} \mid \emptyset \rangle \right\}.$$

In fact,  $\mathfrak{I}^* \subset \mathfrak{I}_m(\mathcal{D})$ , where  $\mathcal{D}$  is the DAG in Figure 10(i). In this case:

$$\mathfrak{I}_{\{t\}}^{\emptyset} = \left\{ \langle \{a, x\}, \{b\} \mid \emptyset \rangle, \langle \{b, y\}, \{a\} \mid \emptyset \rangle \right\}, \mathfrak{I}_{\emptyset}^{\{t\}} = \left\{ \langle \{a, x\}, \{b, y\} \mid \emptyset \rangle \right\}.$$

## 4.2 Marginalizing and Conditioning for Ancestral Graphs

Given an ancestral graph  $\mathcal{G}$  with vertex set  $V$ , for arbitrary disjoint sets  $S$ ,  $L$  (both possibly empty) we now define a transformation:

$$\mathcal{G} \mapsto \mathcal{G}_L^S$$

The main result of this section will be:

**Theorem 4.18** *If  $\mathcal{G}$  is an ancestral graph over  $V$ , and  $S \dot{\cup} L \subset V$ , then*

$$\mathfrak{I}_m(\mathcal{G})_L^S = \mathfrak{I}_m(\mathcal{G}_L^S)$$

(where  $A \dot{\cup} B$  denotes the disjoint union of  $A$  and  $B$ ).

In words, the independence model corresponding to the transformed graph is the independence model obtained by marginalizing and conditioning the independence model of the original graph.

Though we define this transformation for any ancestral graph  $\mathcal{G}$ , our primary motivation is the case in which  $\mathcal{G}$  is a DAG, representing some data generating process that is partially observed (corresponding to marginalization) and where selection effects may be present (corresponding to conditioning). See Cox and Wermuth (1996) for further discussion of data-generating processes, marginalizing and conditioning.

#### 4.2.1 Definition of $\mathcal{G}_L^S$

Graph  $\mathcal{G}_L^S$  has vertex set  $V \setminus (S \cup L)$ , and edges specified as follows:

If  $\alpha, \beta$ , are s.t.  $\forall Z$ , with  $Z \subseteq V \setminus (S \cup L \cup \{\alpha, \beta\})$ ,

$$\langle \{\alpha\}, \{\beta\} | Z \cup S \rangle \notin \mathfrak{I}_m(\mathcal{G}),$$

$$\text{and } \left\{ \begin{array}{l} \alpha \in \text{ant}_{\mathcal{G}}(\{\beta\} \cup S); \beta \in \text{ant}_{\mathcal{G}}(\{\alpha\} \cup S) \\ \alpha \notin \text{ant}_{\mathcal{G}}(\{\beta\} \cup S); \beta \in \text{ant}_{\mathcal{G}}(\{\alpha\} \cup S) \\ \alpha \in \text{ant}_{\mathcal{G}}(\{\beta\} \cup S); \beta \notin \text{ant}_{\mathcal{G}}(\{\alpha\} \cup S) \\ \alpha \notin \text{ant}_{\mathcal{G}}(\{\beta\} \cup S); \beta \notin \text{ant}_{\mathcal{G}}(\{\alpha\} \cup S) \end{array} \right\} \text{ then } \left\{ \begin{array}{l} \alpha - \beta \\ \alpha \leftarrow \beta \\ \alpha \rightarrow \beta \\ \alpha \leftrightarrow \beta \end{array} \right\} \text{ in } \mathcal{G}_L^S.$$

In words,  $\mathcal{G}_L^S$  is a graph containing the vertices that are not in  $S$  or  $L$ . Two vertices  $\alpha, \beta$  are adjacent in  $\mathcal{G}_L^S$  if  $\alpha$  and  $\beta$  are m-connected in  $\mathcal{G}$  given any subset that contains all vertices in  $S$  and no vertices in  $L$ . If  $\alpha$  and  $\beta$  are adjacent in  $\mathcal{G}_L^S$  then there is an arrowhead at  $\alpha$  if and only if  $\alpha$  is not anterior to either  $\beta$  or  $S$  in  $\mathcal{G}$ , and a tail otherwise.

Note that if  $\mathcal{G}$  is not maximal then  $\mathcal{G}_L^S \neq \mathcal{G}$ . (See Corollary 5.2.) We will show in Corollary 4.19 that  $\mathcal{G}_L^S$  is always maximal.

#### 4.2.2 Examples

Consider the DAG,  $\mathcal{D}$ , shown in Figure 10(i). The independence model  $\mathfrak{I}_m(\mathcal{D}) \supset \mathfrak{I}^*$ , given in Section 4.1.1. Suppose that we set  $L = \{t\}$ ,  $S = \emptyset$ .

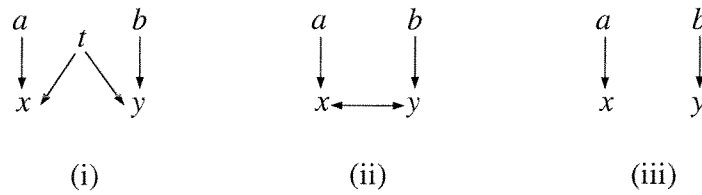


Figure 10: (i) A simple DAG model,  $\mathcal{D}$ ; (ii) the graph  $\mathcal{D}_{\{t\}}^{\emptyset}$ ; (iii) the graph  $\mathcal{D}_{\emptyset}^{\{t\}}$ . See text for further explanation.

First consider the adjacencies that will be present in the transformed graph  $\mathcal{D}_{\{t\}}^{\emptyset}$ . It follows directly from the definition that vertices that are adjacent

in the original graph will also be adjacent in the transformed graph, if they are present in the new graph, since adjacent vertices are  $m$ -connected given any subset of the remaining vertices. Hence the pairs  $(a, x)$  and  $(b, y)$  will be adjacent in  $\mathcal{D}_{\{t\}}^{\emptyset}$ . In addition,  $x$  and  $y$  will be adjacent since any set  $m$ -separating  $x$  and  $y$  in  $\mathcal{D}$  contains  $t$ , hence there is no set  $Z \subseteq \{a, b\}$  such that  $\langle \{x\}, \{y\} \mid Z \rangle \in \mathfrak{I}_m(\mathcal{D})$ . Since  $\langle \{a\}, \{b, y\} \mid \emptyset \rangle, \langle \{b\}, \{a, x\} \mid \emptyset \rangle \in \mathfrak{I}_m(\mathcal{D})$  there are no other adjacencies. It remains to determine the type of these three edges in  $\mathcal{D}_{\{t\}}^{\emptyset}$ . Since  $x \notin \text{ant}_{\mathcal{D}}(y)$ , and  $y \notin \text{ant}_{\mathcal{D}}(x)$ , the edge between  $x$  and  $y$  is of the form  $x \leftrightarrow y$ . Similarly the other edges are  $a \rightarrow x$  and  $b \rightarrow y$ . Thus the graph  $\mathcal{D}_{\{t\}}^{\emptyset}$  is as shown in Figure 10(ii). Observe that  $\mathfrak{I}^*_{\{t\}} \subset \mathfrak{I}_m(\mathcal{D}_{\{t\}}^{\emptyset})$ .

Now suppose that  $L = \emptyset, S = \{t\}$ . Since  $\langle \{a, x\}, \{b, y\} \mid \{t\} \rangle \in \mathfrak{I}_m(\mathcal{D})$ , it follows that  $(a, x)$  and  $(b, y)$  are the only pairs of adjacent vertices present in the transformed graph  $\mathcal{D}_{\emptyset}^{\{t\}}$ , hence this graph takes the form shown in Figure 10(iii). Note that  $\mathfrak{I}^*_{\emptyset}^{\{t\}} \subset \mathfrak{I}_m(\mathcal{D}_{\emptyset}^{\{t\}})$ .

Another example of this transformation is given in Figure 11, with a more complex DAG  $\mathcal{D}'$ . Note the edge between  $a$  and  $c$  that is present in  $\mathcal{D}'^{\{s\}}_{\{l_1, l_2\}}$ .

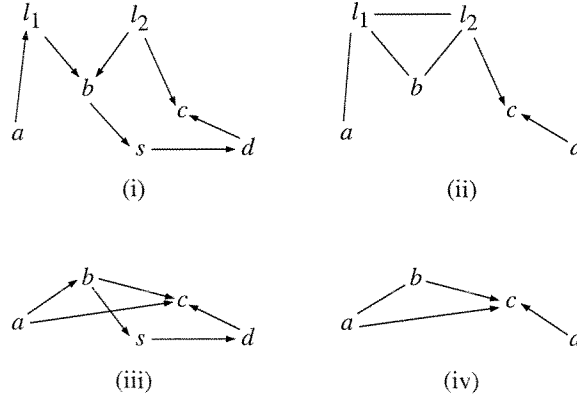


Figure 11: (i) Another DAG,  $\mathcal{D}'$ ; (ii) the graph  $\mathcal{D}'^{\{s\}}_{\emptyset}$ ; (iii) the graph  $\mathcal{D}'^{\emptyset}_{\{l_1, l_2\}}$ ; (iv) the graph  $\mathcal{D}'^{\{s\}}_{\{l_1, l_2\}}$ .

### 4.2.3 Adjacencies in $\mathcal{G}_L^S$ and inducing paths

A path  $\pi$  between  $\alpha$  and  $\beta$  on which every collider is an ancestor of  $\{\alpha, \beta\} \cup S$  and every non-collider is in  $L$ , is called an *inducing path with respect to  $S$  and  $L$* . This is a generalization of the definition introduced by Verma and Pearl (1990). An inducing path with respect to  $S = \emptyset, L = \emptyset$  is called *primitive*. Note that if  $\alpha, \beta \in V \setminus (S \cup L)$ , and  $\alpha, \beta$  are adjacent in  $\mathcal{G}$  then the edge joining  $\alpha$  and  $\beta$  is (trivially) an inducing path w.r.t.  $S$  and  $L$  in  $\mathcal{G}$ .

In Figure 10(i) the path  $x \leftarrow t \rightarrow y$  forms an inducing path w.r.t.  $S = \emptyset, L = \{t\}$ ; in Figure 11(i) the path  $a \rightarrow l_1 \rightarrow b \leftarrow l_2 \rightarrow c$  forms an inducing path w.r.t.  $S = \{s\}, L = \{l_1, l_2\}$ ; in Figure 9(a),  $\gamma \leftrightarrow \beta \leftrightarrow \alpha \leftrightarrow \delta$  forms a primitive inducing path between  $\gamma$  and  $\delta$ . (Other inducing paths are also present in these graphs.)

**Theorem 4.2** *If  $\mathcal{G}$  is an ancestral graph, with vertex set  $V = O \dot{\cup} S \dot{\cup} L$ , and  $\alpha, \beta \in O$  then the following six conditions are equivalent:*

- (i) *There is an edge between  $\alpha$  and  $\beta$  in  $\mathcal{G}_L^S$ .*
- (ii) *There is an inducing path between  $\alpha$  and  $\beta$  w.r.t.  $S$  and  $L$  in  $\mathcal{G}$ .*
- (iii) *There is a path between  $\alpha$  and  $\beta$  in  $(\mathcal{G}_{\text{ant}(\{\alpha, \beta\} \cup S)})^a$  on which every vertex, except the endpoints, is in  $L$ .*
- (iv) *The vertices in  $\text{ant}(\{\alpha, \beta\} \cup S)$  that are not in  $L \cup \{\alpha, \beta\}$  do not m-separate  $\alpha$  and  $\beta$  in  $\mathcal{G}$ :*

$$\left\langle \{\alpha\}, \{\beta\} \middle| \text{ant}(\{\alpha, \beta\} \cup S) \setminus (L \cup \{\alpha, \beta\}) \right\rangle \notin \mathfrak{I}_m(\mathcal{G}).$$

$$(v) \quad \forall Z, Z \subseteq V \setminus (S \cup L \cup \{\alpha, \beta\}), \left\langle \{\alpha\}, \{\beta\} \middle| Z \cup S \right\rangle \notin \mathfrak{I}_m(\mathcal{G}).$$

$$(vi) \quad \forall Z, Z \subseteq V \setminus (S \cup L \cup \{\alpha, \beta\}), \left\langle \{\alpha\}, \{\beta\} \middle| Z \right\rangle \notin \mathfrak{I}_m(\mathcal{G})_L^S.$$

**Proof:** Let  $Z^* = \text{ant}(\{\alpha, \beta\} \cup S) \setminus (L \cup \{\alpha, \beta\})$ . By Proposition 2.2(i)

$$\begin{aligned} \text{ant}(\{\alpha, \beta\} \cup Z^*) &= \text{ant}(\{\alpha, \beta\} \cup (\text{ant}(\{\alpha, \beta\} \cup S) \setminus (L \cup \{\alpha, \beta\}))) \\ &= \text{ant}(\text{ant}(\{\alpha, \beta\} \cup S) \setminus L) \\ &= \text{ant}(\{\alpha, \beta\} \cup S). \end{aligned} \tag{†}$$

In addition, let  $T^* = \text{ant}(\{\alpha, \beta\} \cup S) \cap (L \cup \{\alpha, \beta\})$ , so

$$T^* \dot{\cup} Z^* = \text{ant}(\{\alpha, \beta\} \cup Z^*) \tag{‡}$$

- (iii) $\Leftrightarrow$ (iv) Since, by Theorem 3.18,  $\mathfrak{I}_{m^*}(\mathcal{G}) = \mathfrak{I}_m(\mathcal{G})$ , (iv) holds if and only if there is a path  $\mu$  in  $(\mathcal{G}_{\text{ant}(\{\alpha, \beta\} \cup Z^*)})^a$  on which no vertex is in  $Z^*$ , and hence by  $(\ddagger)$  every vertex is in  $T^*$ . Further, by  $(\dagger)$ ,  $\mathcal{G}_{\text{ant}(\{\alpha, \beta\} \cup Z^*)} = \mathcal{G}_{\text{ant}(\{\alpha, \beta\} \cup S)}$ , hence by the definition of  $T^*$ ,  $\mu$  satisfies the conditions given in (iii).
- (ii) $\Rightarrow$ (iv) If there is an inducing path  $\pi$  in  $\mathcal{G}$  w.r.t.  $S$  and  $L$ , then no non-collider on  $\pi$  is in  $Z^*$ , since  $Z^* \cap L = \emptyset$ , and any collider on  $\pi$  is in  $\text{an}(\{\alpha, \beta\} \cup S) \subseteq \text{ant}(\{\alpha, \beta\} \cup S) = \text{ant}(\{\alpha, \beta\} \cup Z^*)$  by  $(\dagger)$ . Hence by Corollary 3.15 there is a path  $\pi^*$  which m-connects  $\alpha$  and  $\beta$  given  $Z^*$  in  $\mathcal{G}$  as required.
- (iv) $\Rightarrow$ (ii) Let  $\nu$  be a path which m-connects  $\alpha$  and  $\beta$  given  $Z^*$ . By Lemma 3.13 and  $(\dagger)$ , every vertex on  $\nu$  is in  $\text{ant}(\{\alpha, \beta\} \cup S)$ , hence by Lemma 3.2(b) and Corollary 3.3, every collider is in  $\text{an}(\{\alpha, \beta\} \cup S)$ . Every non-collider is in  $\text{ant}(\{\alpha, \beta\} \cup S) \setminus Z^* \subseteq L \cup \{\alpha, \beta\}$ , so every non-collider is in  $L$ . Hence  $\nu$  is an inducing path w.r.t.  $S$  and  $L$  in  $\mathcal{G}$ .
- (iii) $\Rightarrow$ (v) Every edge present in  $(\mathcal{G}_{\text{ant}(\{\alpha, \beta\} \cup S)})^a$  is also present in  $(\mathcal{G}_{\text{ant}(\{\alpha, \beta\} \cup Z \cup S)})^a$ . The implication then follows since every non-endpoint vertex on the path is in  $L$ .
- (v) $\Rightarrow$ (iv) This follows trivially taking  $Z = Z^* \setminus S$ .
- (v) $\Leftrightarrow$ (i) Definition of  $\mathcal{G}_L^S$ .
- (v) $\Leftrightarrow$ (vi) Definition of  $\mathfrak{I}_m(\mathcal{G})_L^S$ . □

An important consequence of condition (iv) in this Theorem is that a single test of m-separation in  $\mathcal{G}$  is sufficient to determine whether or not a given adjacency is present in  $\mathcal{G}_L^S$ ; it is not necessary to test every subset of  $V \setminus (S \cup L \cup \{\alpha, \beta\})$ . Likewise properties (ii) and (iii) provide conditions that can be tested in polynomial time.

#### 4.2.4 Primitive inducing paths and maximality

**Corollary 4.3** *If  $\mathcal{G}$  is an ancestral graph, then there is no set  $Z$ ,  $(\alpha, \beta \notin Z)$ , such that  $\langle \{\alpha\}, \{\beta\} \mid Z \rangle \in \mathfrak{I}_m(\mathcal{G})$  if and only if there is a primitive inducing path between  $\alpha$  and  $\beta$  in  $\mathcal{G}$ .*

**Proof:** The result follows from (ii)  $\Leftrightarrow$  (v) in Theorem 4.2 with  $S = \emptyset, L = \emptyset$ . □

**Corollary 4.4** *Every non-maximal ancestral graph contains a primitive inducing path between a pair of non-adjacent vertices.*

**Proof:** Immediate by the definition of maximality and Corollary 4.3.  $\square$

Primitive inducing paths with more than one edge take a very special form, as described in the next Lemma, and illustrated by the inducing path  $\gamma \leftrightarrow \beta \leftrightarrow \alpha \leftrightarrow \delta$  in Figure 9(a).

**Lemma 4.5** *Let  $\mathcal{G}$  be an ancestral graph. If  $\pi$  is a primitive inducing path between  $\alpha$  and  $\beta$  in  $\mathcal{G}$ , and  $\pi$  contains more than one edge, then:*

- (i) *every non-endpoint vertex on  $\pi$  is a collider and in  $\text{ant}_{\mathcal{G}}(\{\alpha, \beta\})$ ;*
- (ii)  *$\alpha \notin \text{ant}_{\mathcal{G}}(\beta)$  and  $\beta \notin \text{ant}_{\mathcal{G}}(\alpha)$ ;*
- (iii) *every edge on  $\pi$  is bi-directed.*

**Proof:** (i) is a direct consequence of the definition of a primitive inducing path. Consider the vertex  $\gamma$  which is adjacent to  $\alpha$  on  $\pi$ . By (i),  $\gamma$  is a collider on  $\pi$ , so  $\gamma \in \text{sp}_{\mathcal{G}}(\alpha) \cup \text{ch}_{\mathcal{G}}(\alpha)$ , so  $\gamma \notin \text{ant}_{\mathcal{G}}(\alpha)$  as  $\mathcal{G}$  is ancestral. Hence by (i)  $\gamma \in \text{ant}_{\mathcal{G}}(\beta)$ . If  $\beta \in \text{ant}_{\mathcal{G}}(\alpha)$  then  $\gamma \in \text{ant}_{\mathcal{G}}(\alpha)$ , but this is a contradiction. Thus  $\beta \notin \text{ant}_{\mathcal{G}}(\alpha)$ . By a similar argument  $\alpha \notin \text{ant}_{\mathcal{G}}(\beta)$ , establishing (ii). (iii) follows directly from (i) and (ii), since  $\mathcal{G}$  is ancestral.  $\square$

Lemma 4.5 (ii) has the following consequence:

**Corollary 4.6** *In a maximal ancestral graph  $\mathcal{G}$ , if there is a primitive inducing path between  $\alpha$  and  $\beta$  containing more than one edge, then there is an edge  $\alpha \leftrightarrow \beta$  in  $\mathcal{G}$ .*

**Proof:** Since  $\mathcal{G}$  is maximal, by Corollary 4.3,  $\alpha$  and  $\beta$  are adjacent in  $\mathcal{G}$ . By Lemma 4.5(ii),  $\alpha \notin \text{ant}_{\mathcal{G}}(\beta)$  and  $\beta \notin \text{ant}_{\mathcal{G}}(\alpha)$ , hence by Lemma 3.9, it follows that  $\alpha \leftrightarrow \beta$  in  $\mathcal{G}$ .  $\square$

Note that if  $\mathcal{G}$  is a maximal ancestral graph and  $\mathcal{G}'$  is a subgraph formed by removing an undirected or directed edge from  $\mathcal{G}$  then  $\mathcal{G}'$  is also maximal.

#### 4.2.5 Anterior relations in $\mathcal{G}_L^S$

The next Lemma characterizes the vertices anterior to  $\alpha$  in  $\mathcal{G}_L^S$ .

**Lemma 4.7** *For an ancestral graph  $\mathcal{G}$  with vertex set  $V = O \dot{\cup} S \dot{\cup} L$ , if  $\alpha \in O$  then*

$$\text{ant}_{\mathcal{G}}(\alpha) \setminus (\text{ant}_{\mathcal{G}}(S) \cup L) \subseteq \text{ant}_{\mathcal{G}_L^S}(\alpha) \subseteq \text{ant}_{\mathcal{G}}(\{\alpha\} \cup S) \setminus (S \cup L).$$

In words, if  $\alpha, \beta$  are in  $\mathcal{G}_L^S$  and  $\alpha$  is anterior to  $\beta$  but not  $S$  in  $\mathcal{G}$ , then  $\alpha$  is also anterior to  $\beta$  in  $\mathcal{G}_L^S$ . Conversely, if  $\alpha$  is anterior to  $\beta$  in  $\mathcal{G}_L^S$  then  $\alpha$  is anterior to either  $\beta$  or  $S$  in  $\mathcal{G}$ .

**Proof:** Let  $\mu$  be an anterior path from a vertex  $\beta \in \text{ant}_{\mathcal{G}}(\alpha) \setminus (L \cup \text{ant}_{\mathcal{G}}(S))$  to  $\alpha$  in  $\mathcal{G}$ . Note that no vertex on  $\mu$  is in  $S$ . Consider the subsequence  $\langle \beta \equiv \omega_m, \dots, \omega_i, \dots, \omega_1 \equiv \alpha \rangle$  of vertices on  $\mu$  that are in  $V \setminus (S \cup L)$ . Now the subpath  $\mu(\omega_{i+1}, \omega_i)$  is an anterior path on which every vertex except the endpoints is in  $L$ . Hence  $\omega_i$  and  $\omega_{i+1}$  are adjacent in  $\mathcal{G}_L^S$ . Further since  $\omega_{i+1} \in \text{ant}_{\mathcal{G}}(\omega_i)$  it follows that either  $\omega_{i+1} - \omega_i$  or  $\omega_{i+1} \rightarrow \omega_i$ , hence  $\beta \equiv \omega_m \in \text{ant}_{\mathcal{G}_L^S}(\alpha)$ , as required.

To prove the second assertion, let  $\nu \equiv \langle \phi_n, \dots, \phi_1 \equiv \alpha \rangle$  be an anterior path from a vertex  $\phi_n \in \text{ant}_{\mathcal{G}_L^S}(\alpha)$  to  $\alpha$  in  $\mathcal{G}_L^S$ . For  $1 \leq i < n$ , either  $\phi_{i+1} - \phi_i$  or  $\phi_{i+1} \rightarrow \phi_i$  on  $\nu$ . By definition of  $\mathcal{G}_L^S$ , in either case  $\phi_{i+1} \in \text{ant}_{\mathcal{G}}(\{\phi_i\} \cup S) \setminus (S \cup L)$ . Thus  $\phi_n \in \text{ant}_{\mathcal{G}}(\{\alpha\} \cup S) \setminus (S \cup L)$ .  $\square$

Taking  $S = \emptyset$  in Lemma 4.7 we obtain the following:

**Corollary 4.8** *In an ancestral graph  $\mathcal{G} = (V, E)$  if  $\alpha \in V \setminus L$  then  $\text{ant}_{\mathcal{G}}(\alpha) \setminus L = \text{ant}_{\mathcal{G}_L^{\emptyset}}(\alpha)$ .*

#### 4.2.6 The undirected subgraph of $\mathcal{G}_L^S$

**Lemma 4.9** *If  $\mathcal{G}$  is an ancestral graph with vertex set  $V = O \dot{\cup} S \dot{\cup} L$ , then*

$$(\text{un}_{\mathcal{G}} \cup \text{ant}_{\mathcal{G}}(S)) \setminus (S \cup L) \subseteq \text{un}_{\mathcal{G}_L^S}.$$

In words, any vertex in the undirected subgraph of  $\mathcal{G}$  which is also present in  $\mathcal{G}_L^S$  will also be in the undirected subgraph of  $\mathcal{G}_L^S$ . Likewise any vertex

anterior to  $S$  in  $\mathcal{G}$  will be in the undirected component of  $\mathcal{G}_L^S$  if present in this graph.

**Proof:** Suppose for a contradiction that  $\alpha \in (\text{un}_{\mathcal{G}} \cup \text{ant}_{\mathcal{G}}(S)) \setminus (S \cup L)$ , but  $\alpha \notin \text{un}_{\mathcal{G}_L^S}$ . Hence there is a vertex  $\beta$  such that either  $\beta \leftrightarrow \alpha$  or  $\beta \rightarrow \alpha$  in  $\mathcal{G}_L^S$ . In both cases  $\alpha \notin \text{ant}_{\mathcal{G}}(\{\beta\} \cup S)$ . Thus  $\alpha \notin \text{ant}_{\mathcal{G}}(S)$ . Since  $\alpha$  and  $\beta$  are adjacent in  $\mathcal{G}_L^S$  by Theorem 4.2(ii) there is an inducing path  $\pi$  between  $\alpha$  and  $\beta$  w.r.t.  $S$  and  $L$ , hence every vertex on  $\pi$  is in  $\text{ant}_{\mathcal{G}}(\{\alpha, \beta\} \cup S)$ . If there are no colliders on  $\pi$  then since  $\alpha \in \text{un}_{\mathcal{G}}$ ,  $\pi$  is an anterior path from  $\alpha$  to  $\beta$  so  $\alpha \in \text{ant}_{\mathcal{G}}(\beta)$ , which is a contradiction. If there is a collider on  $\pi$  then let  $\gamma$  be the collider on  $\pi$  closest to  $\alpha$ . Now  $\pi(\alpha, \gamma)$  is an anterior path from  $\alpha$  to  $\gamma$  so  $\alpha \in \text{ant}_{\mathcal{G}}(\gamma)$  but  $\gamma \notin \text{un}_{\mathcal{G}}$ , hence by Lemma 3.8(ii),  $\gamma \notin \text{ant}_{\mathcal{G}}(\alpha)$ . Thus  $\gamma \in \text{ant}_{\mathcal{G}}(\{\beta\} \cup S)$ , and thus  $\alpha \in \text{ant}_{\mathcal{G}}(\{\beta\} \cup S)$ , again a contradiction.  $\square$

**Corollary 4.10** *If  $\mathcal{G}$  is an ancestral graph with  $V = O \dot{\cup} S \dot{\cup} L$  and  $\alpha \in O$  then*

$$\text{ant}_{\mathcal{G}}(\alpha) \setminus (S \cup L) \subseteq \text{un}_{\mathcal{G}_L^S} \cup \text{ant}_{\mathcal{G}_L^S}(\alpha).$$

Thus the vertices anterior to  $\alpha \in \mathcal{G}$  that are also in  $\mathcal{G}_L^S$  either remain anterior to  $\alpha \in \mathcal{G}_L^S$ , or are in  $\text{un}_{\mathcal{G}_L^S}$  (or both).

**Proof:**

$$\begin{aligned} (\text{ant}_{\mathcal{G}}(\alpha)) \setminus (S \cup L) &\subseteq \left( \text{ant}_{\mathcal{G}}(\alpha) \setminus (\text{ant}_{\mathcal{G}}(S) \cup L) \right) \dot{\cup} \left( \text{ant}_{\mathcal{G}}(S) \setminus (S \cup L) \right) \\ (*) &\subseteq \text{ant}_{\mathcal{G}_L^S}(\alpha) \cup \text{un}_{\mathcal{G}_L^S} \end{aligned}$$

The step marked (\*) follows from Lemmas 4.7 and 4.9.  $\square$

**Lemma 4.11** *In an ancestral graph  $\mathcal{G}$ , if  $\alpha \in \text{ant}_{\mathcal{G}_L^S}(\beta)$  and  $\alpha \notin \text{un}_{\mathcal{G}_L^S}$  then  $\alpha \in \text{an}_{\mathcal{G}}(\beta)$ , and  $\alpha \notin \text{ant}_{\mathcal{G}}(S)$ .*

**Proof:** If  $\alpha \notin \text{un}_{\mathcal{G}_L^S}$ , but  $\alpha \in V \setminus (S \cup L)$  then by Lemma 4.9,  $\alpha \notin \text{un}_{\mathcal{G}} \cup \text{ant}_{\mathcal{G}}(S)$ . Since  $\alpha \in \text{ant}_{\mathcal{G}_L^S}(\beta)$  it follows from Lemma 4.7 that  $\alpha \in \text{ant}_{\mathcal{G}}(\{\beta\} \cup S)$ . So  $\alpha \in \text{ant}_{\mathcal{G}}(\beta)$ . Further, since  $\alpha \notin \text{un}_{\mathcal{G}}$ , by Lemma 3.8(iii),  $\alpha \in \text{an}_{\mathcal{G}}(\beta)$ .  $\square$

Consequently, if in  $\mathcal{G}_L^S$   $\alpha$  is anterior to  $\beta$  and there is an arrowhead at  $\alpha$  then  $\alpha$  is an ancestor of  $\beta$  in  $\mathcal{G}$ .



#### 4.2.7 $\mathcal{G}_L^S$ is an ancestral graph

**Theorem 4.12** *If  $\mathcal{G}$  is an arbitrary ancestral graph, with vertex set  $V = O \dot{\cup} S \dot{\cup} L$ , then  $\mathcal{G}_L^S$  is an ancestral graph.*

**Proof:** Clearly  $\mathcal{G}_L^S$  is a mixed graph. Suppose for a contradiction that  $\alpha \in \text{ant}_{\mathcal{G}_L^S}(\text{pa}_{\mathcal{G}_L^S}(\alpha) \cup \text{sp}_{\mathcal{G}_L^S}(\alpha))$ . Suppose  $\alpha \in \text{ant}_{\mathcal{G}_L^S}(\beta)$  with  $\beta \in \text{pa}_{\mathcal{G}_L^S}(\alpha) \cup \text{sp}_{\mathcal{G}_L^S}(\alpha)$ . Then by Lemma 4.7,  $\alpha \in \text{ant}_{\mathcal{G}}(\{\beta\} \cup S)$ . However if  $\beta \in \text{pa}_{\mathcal{G}_L^S}(\alpha) \cup \text{sp}_{\mathcal{G}_L^S}(\alpha)$  then  $\alpha \notin \text{ant}_{\mathcal{G}}(\beta \cup S)$  by definition of  $\mathcal{G}_L^S$ , which is a contradiction. Hence  $\mathcal{G}_L^S$  satisfies condition (i) for an ancestral graph.

Now suppose that  $\text{ne}_{\mathcal{G}_L^S}(\alpha) \neq \emptyset$ . Let  $\beta \in \text{ne}_{\mathcal{G}_L^S}(\alpha)$ . Then by the definition of  $\mathcal{G}_L^S$ ,  $\alpha \in \text{ant}_{\mathcal{G}}(\{\beta\} \cup S)$  and  $\beta \in \text{ant}_{\mathcal{G}}(\{\alpha\} \cup S)$ . Thus either  $\alpha \in \text{ant}_{\mathcal{G}}(S)$  or, by Lemma 3.8(ii),  $\alpha \in \text{un}_{\mathcal{G}}$ . It follows by Lemma 4.9 that  $\alpha \in \text{un}_{\mathcal{G}_L^S}$ , hence  $\text{pa}_{\mathcal{G}_L^S}(\alpha) \cup \text{sp}_{\mathcal{G}_L^S}(\alpha) = \emptyset$ . So  $\mathcal{G}_L^S$  satisfies condition (ii) for an ancestral graph.  $\square$

We will show in Section 4.2.10 that  $\mathcal{G}_L^S$  is a maximal ancestral graph.

#### 4.2.8 Introduction of undirected and bi-directed edges

As stated earlier, we are particularly interested in considering the transformation  $\mathcal{G} \mapsto \mathcal{G}_L^S$  in the case where  $\mathcal{G}$  is a DAG, and hence contains no bi-directed or undirected edges. The following results show that the introduction of undirected edges is naturally associated with conditioning, while bi-directed are associated with marginalizing.

**Proposition 4.13** *If  $\mathcal{G}$  is an ancestral graph which contains no undirected edges, then neither does  $\mathcal{G}_L^\emptyset$ .*

**Proof:** If  $\alpha - \beta$  in  $\mathcal{G}_L^\emptyset$  then, by construction,  $\alpha \in \text{ant}_{\mathcal{G}}(\beta)$ ,  $\beta \in \text{ant}_{\mathcal{G}}(\alpha)$ . Hence by Lemma 3.8(ii) there is a path composed of undirected edges which joins  $\alpha$  and  $\beta$  in  $\mathcal{G}$ , which is a contradiction.  $\square$

In particular, if we begin with a DAG, then undirected edges will only be present in the transformed graph if  $S \neq \emptyset$ ; likewise it follows from the next Proposition that bi-directed edges will only be present if  $L \neq \emptyset$ .

**Proposition 4.14** *if  $\mathcal{G}$  is an ancestral graph which contains no bi-directed edges then neither does  $\mathcal{G}_{\emptyset}^S$ .*

**Proof:** If  $\alpha \leftrightarrow \beta$  in  $\mathcal{G}_{\emptyset}^S$  then  $\alpha \notin \text{ant}_{\mathcal{G}}(\{\beta\} \cup S)$  and  $\beta \notin \text{ant}_{\mathcal{G}}(\{\alpha\} \cup S)$ . Since there are no bi-directed edges in  $\mathcal{G}$  it follows that  $\alpha$  and  $\beta$  are not adjacent in  $\mathcal{G}$ . Since  $L = \emptyset$ , it further follows that any inducing path has the form  $\alpha \rightarrow \sigma \leftarrow \beta$ , where  $\sigma \in \text{ant}_{\mathcal{G}}(S)$ , contradicting  $\alpha, \beta \notin \text{ant}_{\mathcal{G}}(S)$ .  $\square$

#### 4.2.9 The independence model $\mathfrak{I}_m(\mathcal{G}_L^S)$

The following Lemmas and Corollary are required to prove Theorem 4.18.

**Lemma 4.15** *If  $\mathcal{G}$  is an ancestral graph with  $V = O \dot{\cup} S \dot{\cup} L$ , and  $\beta \in \text{pa}_{\mathcal{G}_L^S}(\alpha) \cup \text{sp}_{\mathcal{G}_L^S}(\alpha)$  then  $\alpha$  is not anterior to any vertex on an inducing path (w.r.t.  $S$  and  $L$ ) between  $\alpha$  and  $\beta$  in  $\mathcal{G}$ .*

**Proof:** If  $\beta \in \text{pa}_{\mathcal{G}_L^S}(\alpha) \cup \text{sp}_{\mathcal{G}_L^S}(\alpha)$ , then  $\alpha \notin \text{un}_{\mathcal{G}_L^S}$ . It then follows by Lemma 4.9 that  $\alpha \notin \text{un}_{\mathcal{G}}$ , and by construction of  $\mathcal{G}_L^S$  that  $\alpha \notin \text{ant}_{\mathcal{G}}(\{\beta\} \cup S)$ . A vertex  $\gamma$  on an inducing path between  $\alpha$  and  $\beta$  is in  $\text{ant}_{\mathcal{G}}(\{\alpha, \beta\} \cup S)$ . If  $\alpha \in \text{ant}_{\mathcal{G}}(\gamma)$  then by Lemma 3.8(ii)  $\gamma \notin \text{ant}_{\mathcal{G}}(\alpha)$ , since  $\alpha \notin \text{un}_{\mathcal{G}}$ . Thus  $\gamma \in \text{ant}_{\mathcal{G}}(\{\beta\} \cup S)$  but then  $\alpha \in \text{ant}_{\mathcal{G}}(\{\beta\} \cup S)$ , which is a contradiction.  $\square$

**Corollary 4.16** *If  $\alpha \leftrightarrow \beta$  or  $\alpha \leftarrow \beta$  in  $\mathcal{G}_L^S$  and  $\langle \alpha, \phi_1, \dots, \phi_k, \beta \rangle$  is an inducing path (w.r.t.  $S$  and  $L$ ) in  $\mathcal{G}$  then  $\phi_1 \in \text{pa}_{\mathcal{G}}(\alpha) \cup \text{sp}_{\mathcal{G}}(\alpha)$ .*

**Proof:** By Lemma 4.15  $\alpha \notin \text{ant}_{\mathcal{G}}(\phi_1)$ , hence  $\phi_1 \in \text{pa}_{\mathcal{G}}(\alpha) \cup \text{sp}_{\mathcal{G}}(\alpha)$ .  $\square$

The next Lemma forms the core of the proof of Theorem 4.18.

**Lemma 4.17** *If  $\mathcal{G}$  is an ancestral graph with  $V = O \dot{\cup} S \dot{\cup} L$ ,  $Z \dot{\cup} \{\alpha, \beta\} \subseteq O$  then the following are equivalent:*

- (i) *There is an edge between  $\alpha$  and  $\beta$  in  $\left( (\mathcal{G}_L^S)_{\text{ant}_{\mathcal{G}_L^S}(\{\alpha, \beta\} \cup Z)} \right)^a$ .*
- (ii) *There is a path between  $\alpha$  and  $\beta$  in  $\left( \mathcal{G}_{\text{ant}_{\mathcal{G}}(\{\alpha, \beta\} \cup Z \cup S)} \right)^a$  on which every vertex, except the endpoints, is in  $L$ .*
- (iii) *There is a path which  $m$ -connects  $\alpha$  and  $\beta$  in  $\mathcal{G}$  given*

$$\text{ant}_{\mathcal{G}}(\{\alpha, \beta\} \cup Z \cup S) \setminus (L \cup \{\alpha, \beta\}).$$

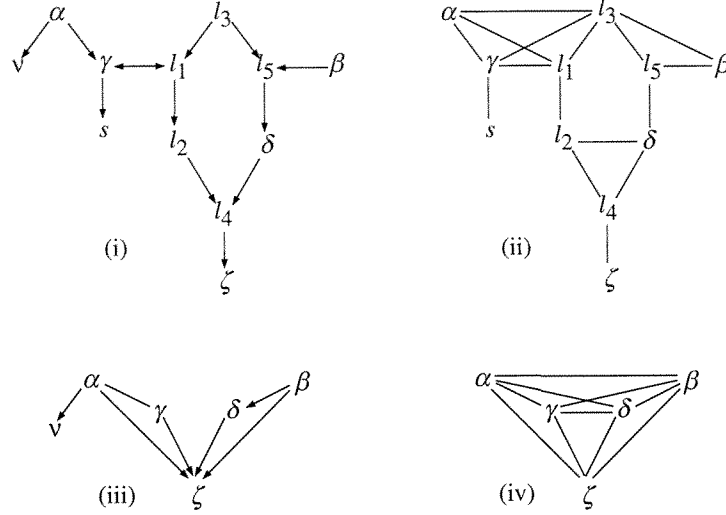


Figure 12: Example of Lemma 4.17: (i) an ancestral graph  $\mathcal{G}$ ; (ii) the augmented graph  $(\mathcal{G}_{\text{ant}_{\mathcal{G}}(\{\alpha, \beta\} \cup Z \cup S)})^a$ ; (iii) the graph  $\mathcal{G}_L^S$ ; (iv) the augmented graph  $((\mathcal{G}_L^S)_{\text{ant}_{\mathcal{G}_L^S}(\{\alpha, \beta\} \cup Z)})^a$ ; (where  $Z = \{\zeta\}$ ,  $S = \{s\}$  and  $L = \{l_1, l_2, l_3, l_4, l_5\}$ ). See text for further explanation.

Figure 12 gives an example of this Lemma, continued below, to illustrate the constructions used in two of the following proofs.

**Proof:**

- (i) $\Rightarrow$ (ii) By (i) there is a path  $\pi$  between  $\alpha$  and  $\beta$  in  $\mathcal{G}_L^S$  on which every non-endpoint vertex is a collider and an ancestor of  $Z \cup \{\alpha, \beta\}$  in  $\mathcal{G}_L^S$ . Let the vertices on  $\pi$  be denoted by  $\langle \omega_0, \dots, \omega_{n+1} \rangle$ , ( $\alpha = \omega_0$ ,  $\beta = \omega_{n+1}$ ). By Lemma 4.7  $\omega_i \in \text{ant}_{\mathcal{G}}(\{\alpha, \beta\} \cup Z \cup S)$ . By Theorem 4.2 there is a path  $\nu_i$  between  $\omega_i$  and  $\omega_{i+1}$  in  $\mathcal{G}_{\text{ant}(\{\omega_i, \omega_{i+1}\} \cup S)}$  on which every non-collider is in  $L$ . The path  $\nu_i$  exists in  $\mathcal{G}_{\text{ant}(\{\alpha, \beta\} \cup Z \cup S)}$  as it is a supergraph of  $\mathcal{G}_{\text{ant}(\{\omega_i, \omega_{i+1}\} \cup S)}$ . Let  $\mathfrak{s}$  be the sequence of vertices formed by concatenating the sequences of vertices on each of the paths  $\nu_i$ . (The same vertex may occur more than once in  $\mathfrak{s}$ .) Let  $\langle \psi_1, \dots, \psi_r \rangle$  be the subsequence of vertices in  $\mathfrak{s}$  each of which is a non-collider on some path  $\nu_i$ , and let  $\psi_0 = \alpha, \psi_{r+1} = \beta$ . Since  $\psi_1, \dots, \psi_r \in L$ , it is

sufficient to show that for  $0 \leq j < r+1$ , if  $\psi_j \neq \psi_{j+1}$  then  $\psi_j - \psi_{j+1}$  in  $(\mathcal{G}_{\text{ant}(\{\alpha, \beta\} \cup Z \cup S)})^a$ . Suppose  $\psi_j \neq \psi_{j+1}$ , there are now two cases:

- (a)  $\psi_j$  and  $\psi_{j+1}$  both occur on the same path  $\nu_i$ . In this case  $\psi_j$  and  $\psi_{j+1}$  are connected in  $(\mathcal{G}_{\text{ant}(\{\alpha, \beta\} \cup Z \cup S)})^a$  by the augmented edge corresponding to the collider path  $\nu_i(\psi_j, \psi_{j+1})$ .
- (b)  $\psi_j$  and  $\psi_{j+1}$  occur on different paths,  $\nu_{i_j}$  and  $\nu_{i_{j+1}}$ . Consider the subsequence  $\mathfrak{s}(\psi_j, \psi_{j+1})$ , denoted by  $\langle \phi_0, \phi_1, \dots, \phi_q, \phi_{q+1} \rangle$ , with  $\phi_0 = \psi_j, \phi_{q+1} = \psi_{j+1}$ . For  $1 \leq k \leq q$  any vertex  $\phi_k$  is either on  $\nu_i$  or is an endpoint  $\omega_i$  of  $\nu_i$  with  $i_j < i \leq i_{j+1}$ . In the former case since  $\psi_j$  and  $\psi_{j+1}$  are consecutive non-colliders in  $\mathfrak{s}$ ,  $\phi_k$  is a collider on  $\nu_i$ . In the latter case by Corollary 4.16,  $\phi_{k-1}, \phi_{k+1} \in \text{pa}_{\mathcal{G}}(\omega_i) \cup \text{sp}_{\mathcal{G}}(\omega_i)$  since  $\omega_i$  is a collider on  $\pi$ . Thus for  $1 \leq k < q$ ,  $\phi_k \leftrightarrow \phi_{k+1}$ , moreover  $\psi_j \rightarrow \phi_1$  or  $\psi_j \leftrightarrow \phi_1$ , and  $\phi_q \leftarrow \psi_{j+1}$  or  $\phi_q \leftrightarrow \psi_{j+1}$ . Hence  $\psi_j$  and  $\psi_{j+1}$  are collider connected in  $\mathcal{G}_{\text{ant}(\{\alpha, \beta\} \cup Z \cup S)$ , and consequently adjacent in  $(\mathcal{G}_{\text{ant}(\{\alpha, \beta\} \cup Z \cup S)})^a$ .  $\square$

Applying the construction in the previous proof to the example in Figure 12, we have  $\pi = \langle \alpha, \zeta, \beta \rangle = \langle \omega_0, \omega_1, \omega_2 \rangle$  in  $\mathcal{G}_L^S$ , hence  $n = 1$ . Further,  $\nu_0 = \langle \alpha, \gamma, l_1, l_2, l_4, \zeta \rangle$  and  $\nu_1 = \langle \zeta, l_4, l_2, l_1, l_3, l_5, \beta \rangle$ , hence  $\mathfrak{s} = \langle \alpha, \gamma, l_1, l_2, l_4, \zeta, l_4, l_2, l_1, l_3, l_5, \beta \rangle$ . Now,  $\langle \psi_0, \dots, \psi_9 \rangle = \langle \alpha, l_1, l_2, l_4, l_4, l_2, l_1, l_3, l_5, \beta \rangle$ , so  $r = 8$ . For  $j \neq 3$ , case (a) applies since  $\psi_j$  and  $\psi_{j+1}$  occur on the same path  $\nu_i$ ; for  $j = 3$ ,  $\psi_j = \psi_{j+1}$ .

(ii) $\Leftrightarrow$ (iii) This follows from Proposition 2.2 together with the definition and equivalence of m-separation and m\*-separation (Theorem 3.18).

(iii) $\Rightarrow$ (i) Let  $Z^* = \text{ant}_{\mathcal{G}}(\{\alpha, \beta\} \cup Z \cup S) \setminus (L \cup \{\alpha, \beta\})$ , and let  $\pi$  be a path which m-connects  $\alpha$  and  $\beta$  given  $Z^*$  in  $\mathcal{G}$ . By Lemma 3.13 every non-collider on  $\pi$  is in  $\text{ant}_{\mathcal{G}}(\{\alpha, \beta\} \cup Z^*) = \text{ant}_{\mathcal{G}}(\{\alpha, \beta\} \cup Z \cup S)$  by Propositions 2.1(iii) and 2.2(i). Every non-collider on  $\pi$  is in  $L$  and every collider is an ancestor of  $Z^*$ . Let  $\langle \psi_1, \dots, \psi_t \rangle$  denote the sequence of colliders on  $\pi$  that are not in  $\text{ant}_{\mathcal{G}}(S)$ , and let  $\psi_0 = \alpha$  and  $\psi_{t+1} = \beta$ . For  $1 \leq i \leq t$  let  $\phi_i$  be the first vertex in  $O$  on a shortest directed path from  $\psi_i$  to a vertex  $\zeta_i \in Z^* \setminus \text{ant}_{\mathcal{G}}(S) \subset \text{ant}_{\mathcal{G}}(Z \cup \{\alpha, \beta\}) \setminus (\text{ant}_{\mathcal{G}}(S) \cup L)$ , denoted  $\nu_i$ . Again let  $\phi_0 = \alpha, \phi_{t+1} = \beta$ . Denote the sequence  $\langle \phi_0, \dots, \phi_{t+1} \rangle$  by  $\mathfrak{t}$ . Finally, let  $\mathfrak{s}$  be a subsequence of  $\mathfrak{t}$  constructed as follows:

$i(0) = 0$ , so  $\phi_{i(0)} = \alpha$ ;

$i(k+1)$  is the greatest  $j > i(k)$  with  $\{\phi_{i(k)}, \dots, \phi_j\} \subseteq \text{ant}_{\mathcal{G}}(\{\phi_{i(k)}, \phi_j\})$ .

Note that if  $i(k) < t$  then  $i(k+1)$  is guaranteed to exist since

$$\{\phi_{i(k)}, \phi_{i(k)+1}\} \subseteq \text{ant}_{\mathcal{G}}(\{\phi_{i(k)}, \phi_{i(k)+1}\}).$$

In addition, the vertices in  $\mathfrak{s}$  are distinct. Let  $s$  be such that  $i(s+1) = t+1$ , so  $\phi_{i(s+1)} = \beta$ .

We now show that there is a path connecting  $\phi_{i(k)}$  and  $\phi_{i(k+1)}$  in  $(\mathcal{G}_{\text{ant}_{\mathcal{G}}(\{\phi_{i(k)}, \phi_{i(k+1)}\} \cup S)})^a$  on which every vertex except the endpoints is in  $L$ :  $\phi_{i(k)}$  and  $\psi_{i(k)}$  are connected by the path corresponding to  $\nu_{i(k)}$  in  $(\mathcal{G}_{\text{ant}_{\mathcal{G}}(\{\phi_{i(k)}, \phi_{i(k+1)}\} \cup S)})^a$ , and likewise  $\phi_{i(k+1)}$  and  $\psi_{i(k+1)}$  are connected by the path corresponding to  $\nu_{i(k+1)}$ . In addition, excepting the endpoints  $\phi_{i(k)}$  and  $\phi_{i(k+1)}$ , every vertex on  $\nu_{i(k)}$  and  $\nu_{i(k+1)}$  is in  $L$ . By construction, every collider on  $\pi(\psi_{i(k)}, \psi_{i(k+1)})$  is either in  $\text{ant}_{\mathcal{G}}(\{\phi_{i(k)}, \phi_{i(k+1)}\})$  or  $\text{ant}_{\mathcal{G}}(S)$ . Further, every non-collider  $\gamma$  on  $\pi(\psi_{i(k)}, \psi_{i(k+1)})$  is either anterior to  $\psi_j$  ( $i(k) \leq j \leq i(k+1)$ ) or is anterior to a collider that is in  $\text{ant}_{\mathcal{G}}(S)$ . Thus every vertex on  $\pi(\psi_{i(k)}, \psi_{i(k+1)})$  is in  $\text{ant}_{\mathcal{G}}(\{\phi_{i(k)}, \phi_{i(k+1)}\} \cup S)$ , so this path exists in  $\mathcal{G}_{\text{ant}_{\mathcal{G}}(\{\phi_{i(k)}, \phi_{i(k+1)}\} \cup S)}$ . The sequence of non-colliders on  $\pi(\psi_{i(k)}, \psi_{i(k+1)})$ , all of which are in  $L$ , connect  $\psi_{i(k)}$  and  $\psi_{i(k+1)}$  in  $(\mathcal{G}_{\text{ant}_{\mathcal{G}}(\{\phi_{i(k)}, \phi_{i(k+1)}\} \cup S)})^a$ . It now follows from Theorem 4.2 (iii)  $\Leftrightarrow$  (i) that  $\phi_{i(k)}$  and  $\phi_{i(k+1)}$  are adjacent in  $\mathcal{G}_L^S$ .

Next we show that  $\phi_0 \rightarrow \phi_{i(1)}$  or  $\phi_0 \leftrightarrow \phi_{i(1)}$ ,  $\phi_{i(s)} \leftarrow \phi_{i(s+1)}$  or  $\phi_{i(s)} \leftrightarrow \phi_{i(s+1)}$  and  $1 \leq k < s$ ,  $\phi_{i(k)} \leftrightarrow \phi_{i(k+1)}$  in  $\mathcal{G}_L^S$ , from which it follows that  $\alpha$  and  $\beta$  are collider connected as required. By construction  $\{\phi_{i(k-1)}, \dots, \phi_{i(k)}\} \subseteq \text{ant}_{\mathcal{G}}(\{\phi_{i(k-1)}, \phi_{i(k)}\})$ , hence if  $\phi_{i(k)} \in \text{ant}_{\mathcal{G}}(\{\phi_{i(k-1)}\})$  then  $\{\phi_{i(k-1)}, \dots, \phi_{i(k)}, \phi_{i(k)+1}\} \subseteq \text{ant}_{\mathcal{G}}(\{\phi_{i(k-1)}, \phi_{i(k)+1}\})$ , and thus  $i(k)$  is not the greatest  $j$  such that

$$\{\phi_{i(k-1)}, \dots, \phi_j\} \subseteq \text{ant}_{\mathcal{G}}(\{\phi_{i(k-1)}, \phi_j\}).$$

Thus  $\phi_{i(k)} \notin \text{ant}_{\mathcal{G}}(\{\phi_{i(k-1)}\})$ , ( $1 \leq k \leq s$ ). Further, since

$$\{\phi_{i(k)}, \dots, \phi_{i(k+1)}\} \subseteq \text{ant}_{\mathcal{G}}(\{\phi_{i(k)}, \phi_{i(k+1)}\}),$$

if  $\phi_{i(k)} \in \text{ant}_{\mathcal{G}}(\{\phi_{i(k+1)}\})$  then

$$\{\phi_{i(k-1)}, \dots, \phi_{i(k+1)}\} \subseteq \text{ant}_{\mathcal{G}}(\{\phi_{i(k-1)}, \phi_{i(k+1)}\}),$$

but in that case  $\phi_{i(k)}$  is not the last such vertex after  $\phi_{i(k-1)}$  in  $\mathbf{t}$ , which is a contradiction. By construction,  $\psi_{i(k)} \in \text{ant}_{\mathcal{G}}(\phi_{i(k)})$  for  $1 \leq k \leq s$ , and  $\psi_{i(k)} \notin \text{ant}_{\mathcal{G}}(S)$ , so  $\phi_{i(k)} \notin \text{ant}_{\mathcal{G}}(S)$ . We have now shown that  $\phi_{i(k)} \notin \text{ant}_{\mathcal{G}}(\{\phi_{i(k-1)}, \phi_{i(k+1)}\} \cup S)$ , for  $1 \leq k \leq s$ . The required orientations now follow from the definition of  $\mathcal{G}_L^S$ .

Finally, since  $\{\phi_{i(1)}, \dots, \phi_{i(s)}\} \subseteq \text{ant}_{\mathcal{G}}(Z \cup \{\alpha, \beta\}) \setminus (\text{ant}_{\mathcal{G}}(S) \cup L)$ , it follows by Lemma 4.7 that  $\{\phi_{i(1)}, \dots, \phi_{i(s)}\} \subseteq \text{ant}_{\mathcal{G}_L^S}(Z \cup \{\alpha, \beta\})$ . Hence every vertex in the sequence  $\mathbf{s}$  occurs in  $(\mathcal{G}_L^S)_{\text{ant}_{\mathcal{G}_L^S}(\{\alpha, \beta\} \cup Z)}$ , and thus  $\alpha$  and  $\beta$  are collider connected in this graph, as required.  $\square$

We now apply the construction in the previous proof to the example in Figure 12. The path  $\pi = \langle \alpha, \gamma, l_1, l_3, l_5, \beta \rangle$  m-connects  $\alpha$  and  $\beta$  given  $Z^* = \text{ant}_{\mathcal{G}}(\{\alpha, \beta\} \cup Z \cup S) \setminus (L \cup \{\alpha, \beta\}) = \{\gamma, \delta, s, \zeta\}$ . It follows that  $\langle \psi_0, \psi_1, \psi_2, \psi_3 \rangle = \langle \alpha, l_1, l_5, \beta \rangle$ , so  $t=2$ ;  $\mathbf{t} = \langle \phi_0, \phi_1, \phi_2, \phi_3 \rangle = \langle \alpha, \zeta, \delta, \beta \rangle$ ,  $\nu_1 = \langle l_1, l_2, l_4, \zeta \rangle$ , and  $\nu_2 = \langle l_5, \delta \rangle$ . It then follows that  $\mathbf{s} = \langle \phi_{i(0)}, \phi_{i(1)}, \phi_{i(2)} \rangle = \langle \alpha, \zeta, \beta \rangle$ , so  $s=1$ . For  $k=0, 1$  the graph  $(\mathcal{G}_{\text{ant}_{\mathcal{G}}(\{\phi_{i(k)}, \phi_{i(k+1)}\} \cup S)})^a$  is the graph shown in Figure 12(ii). Finally, note that  $\mathbf{t}$  does not constitute a collider path between  $\alpha$  and  $\beta$  in  $\mathcal{G}_L^S$ , though the subsequence  $\mathbf{s}$  does, as proved.

We are now ready to prove the main result of this section:

**Theorem 4.18** *If  $\mathcal{G}$  is an ancestral graph over  $V$ , and  $S \dot{\cup} L \subset V$ , then*

$$\mathfrak{I}_m(\mathcal{G})_L^S = \mathfrak{I}_m(\mathcal{G}_L^S)$$

**Proof:** Let  $X \dot{\cup} Y \dot{\cup} Z \subseteq O$ . We now argue as follows:

$$\begin{aligned} \langle X, Y \mid Z \rangle &\notin \mathfrak{I}_m(\mathcal{G})_L^S \\ \Leftrightarrow \langle X, Y \mid Z \cup S \rangle &\notin \mathfrak{I}_m(\mathcal{G}) \\ \Leftrightarrow \text{For some } \alpha \in X, \beta \in Y \text{ there is a path } \pi &\text{ connecting } \alpha \text{ and } \beta \\ &\text{ in } (\mathcal{G}_{\text{ant}_{\mathcal{G}}(\{\alpha, \beta\} \cup Z \cup S)})^a, \text{ on which no vertex is in } Z \cup S. \\ (*) \Leftrightarrow \text{For some } \alpha \in X, \beta \in Y \text{ there is a path } \mu &\text{ connecting } \alpha \text{ and } \beta \\ &\text{ in } ((\mathcal{G}_L^S)_{\text{ant}_{\mathcal{G}_L^S}(\{\alpha, \beta\} \cup Z)})^a \text{ on which no vertex is in } Z. \\ \Leftrightarrow \langle X, Y \mid Z \rangle &\notin \mathfrak{I}_m(\mathcal{G}_L^S) \end{aligned}$$

The equivalence (\*) is justified thus:

Let the subsequence of vertices on  $\pi$  that are in  $O$  be denoted  $\langle \omega_1, \dots, \omega_n \rangle$ . Since  $\omega_i, \omega_{i+1} \in \text{ant}_{\mathcal{G}}(\{\alpha, \beta\} \cup Z \cup S)$ ,

$$(\mathcal{G}_{\text{ant}_{\mathcal{G}}(\{\alpha, \beta\} \cup Z \cup S)})^a = (\mathcal{G}_{\text{ant}_{\mathcal{G}}(\{\omega_i, \omega_{i+1}\} \cup (\{\alpha, \beta\} \cup Z) \cup S)})^a.$$

By Lemma 4.17,  $\omega_i$  and  $\omega_{i+1}$  are adjacent in

$$((\mathcal{G}_L^S)_{\text{ant}_{\mathcal{G}_L^S}(\{\omega_i, \omega_{i+1}\} \cup (\{\alpha, \beta\} \cup Z))})^a,$$

since any vertices occurring between  $\omega_i$  and  $\omega_{i+1}$  on  $\pi$  are in  $L$ . We now show by induction that for  $1 \leq i \leq n$ ,  $\omega_i \in \text{ant}_{\mathcal{G}_L^S}(\{\alpha, \beta\} \cup Z)$ . Since  $\omega_1 = \alpha$ , the claim holds trivially for  $i = 1$ . Now suppose that  $\omega_i \in \text{ant}_{\mathcal{G}_L^S}(\{\alpha, \beta\} \cup Z)$ . If  $\omega_{i+1} \notin \text{ant}_{\mathcal{G}}(S)$  then by Lemma 4.7  $\omega_{i+1} \in \text{ant}_{\mathcal{G}_L^S}(\{\alpha, \beta\} \cup Z)$ . On the other hand, if  $\omega_{i+1} \in \text{ant}_{\mathcal{G}}(S)$  then by Lemma 4.9,  $\omega_{i+1} \in \text{un}_{\mathcal{G}_L^S}$ . It follows that in  $\mathcal{G}_L^S$  either  $\omega_{i+1} - \omega_i$ ,  $\omega_{i+1} \rightarrow \omega_i$ , or  $\omega_{i+1} \rightarrow \gamma$ , where  $\gamma$  is a vertex on a collider path between  $\omega_i$  and  $\omega_{i+1}$  in  $(\mathcal{G}_L^S)_{\text{ant}_{\mathcal{G}_L^S}(\{\omega_i, \omega_{i+1}\} \cup (\{\alpha, \beta\} \cup Z))}$ . Consequently,  $\omega_{i+1} \in \text{ant}_{\mathcal{G}_L^S}(\{\omega_i, \alpha, \beta\} \cup Z) = \text{ant}_{\mathcal{G}_L^S}(\{\alpha, \beta\} \cup Z)$ , by the induction hypothesis. It now follows that for  $1 \leq i \leq n$ ,  $\omega_i$  and  $\omega_{i+1}$  are adjacent in

$$((\mathcal{G}_L^S)_{\text{ant}_{\mathcal{G}_L^S}(\{\alpha, \beta\} \cup Z)})^a = ((\mathcal{G}_L^S)_{\text{ant}_{\mathcal{G}_L^S}(\{\omega_i, \omega_{i+1}\} \cup (\{\alpha, \beta\} \cup Z))})^a,$$

hence  $\alpha$  and  $\beta$  are connected in this graph by a path on which no vertex is in  $Z$ .

Conversely, suppose that the vertices on  $\mu$  are  $\langle v_1, \dots, v_m \rangle$ . Since  $v_j, v_{j+1} \in \text{ant}_{\mathcal{G}_L^S}(\{\alpha, \beta\} \cup Z)$ , by Lemma 4.7  $v_j, v_{j+1} \in \text{ant}_{\mathcal{G}}(\{\alpha, \beta\} \cup Z \cup S)$ . As  $v_j$  and  $v_{j+1}$  are adjacent in

$$((\mathcal{G}_L^S)_{\text{ant}_{\mathcal{G}_L^S}(\{\alpha, \beta\} \cup Z)})^a = ((\mathcal{G}_L^S)_{\text{ant}_{\mathcal{G}_L^S}(\{v_i, v_{i+1}\} \cup (\{\alpha, \beta\} \cup Z))})^a,$$

it follows by Lemma 4.17 that  $v_j$  and  $v_{j+1}$  are connected by a path  $\nu_j$  in

$$(\mathcal{G}_{\text{ant}_{\mathcal{G}}(\{v_i, v_{i+1}\} \cup (\{\alpha, \beta\} \cup Z) \cup S)})^a = (\mathcal{G}_{\text{ant}_{\mathcal{G}}(\{\alpha, \beta\} \cup Z \cup S)})^a$$

on which no vertex is in  $Z \cup S$ . Hence  $\alpha$  and  $\beta$  are also connected by such a path.  $\square$

#### 4.2.10 $\mathcal{G}_L^S$ is a maximal ancestral graph

**Corollary 4.19** *If  $\mathcal{G}$  is an ancestral graph with vertex set  $V = O \dot{\cup} S \dot{\cup} L$  then  $\mathcal{G}_L^S$  is a maximal ancestral graph.*

**Proof:** By definition there is an edge between  $\alpha$  and  $\beta$  in  $\mathcal{G}_L^S$  if and only if for all sets  $Z \subseteq O \setminus \{\alpha, \beta\}$ ,  $\langle \{\alpha\}, \{\beta\} \mid Z \cup S \rangle \notin \mathfrak{I}_m(\mathcal{G})$ , or equivalently  $\langle \{\alpha\}, \{\beta\} \mid Z \rangle \notin \mathfrak{I}_m(\mathcal{G})_L^S$ . Hence by Theorem 4.18, there is an edge between  $\alpha$  and  $\beta$  in  $\mathcal{G}_L^S$  if and only if for all sets  $Z \subseteq O \setminus \{\alpha, \beta\}$ ,  $\langle \{\alpha\}, \{\beta\} \mid Z \rangle \notin \mathfrak{I}_m(\mathcal{G}_L^S)$ . Hence  $\mathcal{G}_L^S$  is maximal.  $\square$

#### 4.2.11 Commutativity

**Theorem 4.20** *If  $\mathcal{G}$  is an ancestral graph with vertex set  $V$ , and  $S_1, S_2, L_1, L_2$  are disjoint subsets of  $V$ , then  $\mathcal{G}_{L_1 \cup L_2}^{S_1 \cup S_2} = (\mathcal{G}_{L_1}^{S_1})_{L_2}^{S_2}$ . Hence the following diagram commutes:*

$$\begin{array}{ccc} \mathcal{G} & \longrightarrow & \mathcal{G}_{L_2}^{S_2} \\ \downarrow & & \downarrow \\ \mathcal{G}_{L_1}^{S_1} & \longrightarrow & \mathcal{G}_{L_1 \cup L_2}^{S_1 \cup S_2} \end{array}$$

Figure 11 gives an example of this Theorem.

**Proof:** We first show that  $\mathcal{G}_{L_1 \cup L_2}^{S_1 \cup S_2}$  and  $(\mathcal{G}_{L_1}^{S_1})_{L_2}^{S_2}$  have the same adjacencies. Let  $\alpha, \beta$  be vertices in  $V \setminus (S_1 \cup S_2 \cup L_1 \cup L_2)$ .

There is an edge between  $\alpha$  and  $\beta$  in  $\mathcal{G}_{L_1 \cup L_2}^{S_1 \cup S_2}$

$$\Leftrightarrow \forall Z \subseteq V \setminus ((S_1 \cup S_2) \cup (L_1 \cup L_2) \cup \{\alpha, \beta\}), \langle \{\alpha\}, \{\beta\} \mid Z \cup (S_1 \cup S_2) \rangle \notin \mathfrak{I}_m(\mathcal{G})$$

$$\Leftrightarrow \forall Z \subseteq (V \setminus (S_1 \cup L_1)) \setminus (S_2 \cup L_2 \cup \{\alpha, \beta\}), \langle \{\alpha\}, \{\beta\} \mid Z \cup S_2 \rangle \notin \mathfrak{I}_m(\mathcal{G})_{L_1}^{S_1}$$

$$(*) \Leftrightarrow \forall Z \subseteq (V \setminus (S_1 \cup L_1)) \setminus (S_2 \cup L_2 \cup \{\alpha, \beta\}), \langle \{\alpha\}, \{\beta\} \mid Z \cup S_2 \rangle \notin \mathfrak{I}_m(\mathcal{G}_{L_1}^{S_1})$$

$$\Leftrightarrow \text{There is an edge between } \alpha \text{ and } \beta \text{ in } (\mathcal{G}_{L_1}^{S_1})_{L_2}^{S_2}.$$

The equivalence marked (\*) follows from Theorem 4.18. Now suppose that  $\alpha$  and  $\beta$  are adjacent in  $\mathcal{G}_{L_1 \cup L_2}^{S_1 \cup S_2}$  and  $(\mathcal{G}_{L_1}^{S_1})_{L_2}^{S_2}$ .



$$\begin{aligned}
\alpha \in \text{ant}_{\mathcal{G}_{[L_1 \cup L_2]}^{S_1 \cup S_2}}(\beta) & \\
\Rightarrow \alpha \in \text{ant}_{\mathcal{G}}(\{\beta\} \cup S_1 \cup S_2) & \quad \text{by Lemma 4.7;} \\
\Rightarrow \alpha \in \text{ant}_{\mathcal{G}_{[L_1]}^{S_1}}(\{\beta\} \cup S_2) \text{ or } \alpha \in \text{un}_{\mathcal{G}_{[L_1]}^{S_1}} & \quad \text{by Corollary 4.10 and Lemma 4.9;} \\
\Rightarrow \alpha \in \text{ant}_{(\mathcal{G}_{[L_1]}^{S_1})_{[L_2]}^{S_2}}(\beta) \text{ or } \alpha \in \text{un}_{(\mathcal{G}_{[L_1]}^{S_1})_{[L_2]}^{S_2}} & \quad \text{by Corollary 4.10 and Lemma 4.9;} \\
\Rightarrow \alpha \in \text{ant}_{(\mathcal{G}_{[L_1]}^{S_1})_{[L_2]}^{S_2}}(\beta) & \quad \text{since } \alpha \text{ and } \beta \text{ are adjacent.}
\end{aligned}$$

Arguing in the other direction,

$$\begin{aligned}
\alpha \in \text{ant}_{(\mathcal{G}_{[L_1]}^{S_1})_{[L_2]}^{S_2}}(\beta) & \Rightarrow \alpha \in \text{ant}_{\mathcal{G}_{[L_1]}^{S_1}}(\{\beta\} \cup S_2) & \text{by Lemma 4.7;} \\
& \Rightarrow \alpha \in \text{ant}_{\mathcal{G}}(\{\beta\} \cup S_1 \cup S_2) & \text{by Lemma 4.7;} \\
& \Rightarrow \alpha \in \text{ant}_{\mathcal{G}_{[L_1 \cup L_2]}^{S_1 \cup S_2}}(\beta) \text{ or } \alpha \in \text{un}_{\mathcal{G}_{[L_1 \cup L_2]}^{S_1 \cup S_2}} & \text{by Corollary 4.10 and Lemma 4.9;} \\
& \Rightarrow \alpha \in \text{ant}_{\mathcal{G}_{[L_1 \cup L_2]}^{S_1 \cup S_2}}(\beta) & \text{since } \alpha \text{ and } \beta \text{ are adjacent.}
\end{aligned}$$

It then follows from Corollary 3.10 that  $\mathcal{G}_{[L_1 \cup L_2]}^{S_1 \cup S_2} = (\mathcal{G}_{[L_1]}^{S_1})_{[L_2]}^{S_2}$  as required.  $\square$

## 5 Extending an Ancestral Graph

In this section we prove two extension results. We first show that every ancestral graph can be extended to a maximal ancestral graph, as stated in Section 3.7. We then show that every maximal ancestral graph may be extended to a complete ancestral graph, and that the edge additions may be ordered so that all the intermediate graphs are also maximal. This latter result parallels well known results for decomposable undirected graphs (see Lauritzen (1996), p.20).

## 5.1 Extension of an Ancestral Graph to a Maximal Ancestral Graph

**Theorem 5.1** *If  $\mathcal{G}$  is an ancestral graph then there exists a unique maximal ancestral graph  $\bar{\mathcal{G}}$  formed by adding bi-directed edges to  $\mathcal{G}$  such that  $\mathfrak{I}_m(\mathcal{G}) = \mathfrak{I}_m(\bar{\mathcal{G}})$ .*

Figure 13 gives a simple example of this Theorem.

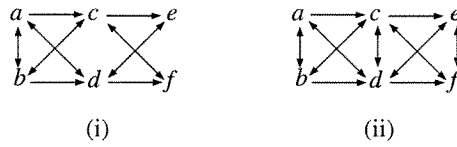


Figure 13: (i) A non-maximal ancestral graph  $\mathcal{G}$ ; (ii) the maximal extension  $\bar{\mathcal{G}}$ . (Every pair of non-adjacent vertices in  $\bar{\mathcal{G}}$  are m-separated either by  $\{c\}$  or  $\{d\}$ .)

**Proof:** Let  $\bar{\mathcal{G}} = \mathcal{G}[\emptyset]$ . It follows from Theorem 4.18 and Proposition 4.1 (i) that

$$\mathfrak{I}_m(\bar{\mathcal{G}}) = \mathfrak{I}_m(\mathcal{G}[\emptyset]) = \mathfrak{I}_m(\mathcal{G})[\emptyset] = \mathfrak{I}_m(\mathcal{G})$$

as required. If  $\alpha$  and  $\beta$  are adjacent in  $\mathcal{G}$  then trivially there is a path m-connecting  $\alpha$  and  $\beta$  given any set  $Z \subset V \setminus \{\alpha, \beta\}$ , hence there is an edge between  $\alpha$  and  $\beta$  in  $\mathcal{G}[\emptyset]$ . Now, by Corollary 4.8,  $\text{ant}_{\mathcal{G}}(\alpha) = \text{ant}_{\mathcal{G}[\emptyset]}(\alpha)$ . Hence by Lemma 3.9 every edge in  $\mathcal{G}$  is inherited by  $\bar{\mathcal{G}} = \mathcal{G}[\emptyset]$ . By Corollary 4.19  $\mathcal{G}[\emptyset]$  is maximal. This establishes the existence of a maximal extension of  $\mathcal{G}$ .

Let  $\bar{\mathcal{G}}$  be a maximal supergraph of  $\mathcal{G}$ . Suppose  $\alpha$  and  $\beta$  are adjacent in  $\bar{\mathcal{G}}$  but are not adjacent in  $\mathcal{G}$ . By Corollary 4.3 there is a primitive inducing path  $\pi$  between  $\alpha$  and  $\beta$  in  $\mathcal{G}$ , containing more than one edge. Since  $\pi$  is present in  $\bar{\mathcal{G}}$ , and this graph is maximal, it follows by Corollary 4.6 that  $\alpha \leftrightarrow \beta$  in  $\bar{\mathcal{G}}$ , as required. This also establishes uniqueness of  $\bar{\mathcal{G}}$ .  $\square$

Three Corollaries are consequences of this result:

**Corollary 5.2**  *$\mathcal{G}$  is a maximal ancestral graph if and only if  $\mathcal{G} = \mathcal{G}[\emptyset]$ .*

**Proof:** Follows directly from the definition of  $\mathcal{G}_{\emptyset}^{\emptyset}$  and Theorem 5.1.  $\square$

The next Corollary establishes the Pairwise Markov property referred to in Section 3.7.

**Corollary 5.3** *If  $\mathcal{G}$  is a maximal ancestral graph and  $\alpha, \beta$  are not adjacent in  $\mathcal{G}$ , then  $\langle \{\alpha\}, \{\beta\} \mid \text{ant}_{\mathcal{G}}(\{\alpha, \beta\}) \setminus \{\alpha, \beta\} \rangle \in \mathfrak{I}_m(\mathcal{G})$ .*

**Proof:** By Corollary 5.2,  $\mathcal{G} = \mathcal{G}_{\emptyset}^{\emptyset}$ . The result then follows by contraposition from Theorem 4.2, properties (i) and (iv).  $\square$

**Corollary 5.4** *If  $\mathcal{G}$  is an ancestral graph,  $\alpha \in \text{ant}_{\mathcal{G}}(\beta)$ , and  $\alpha, \beta$  are not adjacent in  $\mathcal{G}$  then  $\langle \{\alpha\}, \{\beta\} \mid \text{ant}_{\mathcal{G}}(\{\alpha, \beta\}) \setminus \{\alpha, \beta\} \rangle \in \mathfrak{I}_m(\mathcal{G})$ .*

**Proof:** If  $\alpha \in \text{ant}_{\mathcal{G}}(\beta)$  then by Corollary 4.8,  $\alpha \in \text{ant}_{\mathcal{G}_{\emptyset}^{\emptyset}}(\beta)$ . Hence there is no edge  $\alpha \leftrightarrow \beta$  in  $\mathcal{G}_{\emptyset}^{\emptyset}$ , since by Theorem 4.12,  $\mathcal{G}_{\emptyset}^{\emptyset}$  is ancestral. It follows from Theorem 5.1 that  $\alpha$  and  $\beta$  are not adjacent in  $\mathcal{G}_{\emptyset}^{\emptyset}$ . The conclusion then follows from Corollary 5.3.  $\square$

## 5.2 Extension of a Maximal Ancestral Graph to a Complete Graph

For an ancestral graph  $\mathcal{G} = (V, E)$ , the associated complete graph, denoted  $\tilde{\mathcal{G}}$ , is defined as follows:

$\tilde{\mathcal{G}}$  has vertex set  $V$  and an edge between every pair of distinct vertices  $\alpha, \beta$ , specified as follows:

$$\begin{aligned} \alpha - \beta & \quad \text{if } \alpha, \beta \in \text{un}_{\mathcal{G}}, \\ \alpha \rightarrow \beta & \quad \text{if } \alpha \in \text{un}_{\mathcal{G}} \cup \text{ant}_{\mathcal{G}}(\beta) \text{ and } \beta \notin \text{un}_{\mathcal{G}}, \\ \alpha \leftrightarrow \beta & \quad \text{otherwise.} \end{aligned}$$

Thus between each pair of distinct vertices in  $\tilde{\mathcal{G}}$  there will be exactly one edge. Note that although  $\tilde{\mathcal{G}}$  is unique as defined, in general there will be other complete ancestral graphs of which a given graph  $\mathcal{G}$  is a subgraph.

**Lemma 5.5** *If  $\mathcal{G} = (V, E)$  is an ancestral graph, then: (i)  $\mathcal{G}$  is a subgraph of  $\tilde{\mathcal{G}}$ ; (ii)  $\text{un}_{\tilde{\mathcal{G}}} = \text{un}_{\mathcal{G}}$ ; (iii) for all  $\nu \in V$ ,  $\text{ant}_{\tilde{\mathcal{G}}}(\nu) = \text{ant}_{\mathcal{G}}(\nu) \cup \text{un}_{\mathcal{G}}$ ; (iv)  $\tilde{\mathcal{G}}$  is an ancestral graph.*

**Proof:** (i) This follows from the construction of  $\tilde{\mathcal{G}}$ , Lemma 3.7, and  $\text{pa}_{\mathcal{G}}(\nu) \subseteq \text{ant}_{\mathcal{G}}(\nu)$ .

(ii) By construction, if  $\alpha \in \text{un}_{\mathcal{G}}$  then  $\text{pa}_{\tilde{\mathcal{G}}}(\alpha) \cup \text{sp}_{\tilde{\mathcal{G}}}(\alpha) = \emptyset$  hence  $\alpha \in \text{un}_{\tilde{\mathcal{G}}}$ . Conversely, if  $\alpha \notin \text{un}_{\mathcal{G}}$  then  $\text{pa}_{\mathcal{G}}(\alpha) \cup \text{sp}_{\mathcal{G}}(\alpha) \neq \emptyset$ . By (i),  $\text{pa}_{\tilde{\mathcal{G}}}(\alpha) \cup \text{sp}_{\tilde{\mathcal{G}}}(\alpha) \neq \emptyset$ , so  $\alpha \notin \text{un}_{\tilde{\mathcal{G}}}$ . Thus  $\text{un}_{\tilde{\mathcal{G}}} = \text{un}_{\mathcal{G}}$  as required.

(iii) By (i),  $\text{ant}_{\mathcal{G}}(\nu) \subseteq \text{ant}_{\tilde{\mathcal{G}}}(\nu)$ , further, by construction,  $\text{un}_{\tilde{\mathcal{G}}} \subseteq \text{ant}_{\tilde{\mathcal{G}}}(\nu)$ , thus  $\text{ant}_{\mathcal{G}}(\nu) \cup \text{un}_{\mathcal{G}} \subseteq \text{ant}_{\tilde{\mathcal{G}}}(\nu)$ . Conversely, if  $\alpha \in \text{ant}_{\tilde{\mathcal{G}}}(\nu_0)$  then either  $\alpha \in \text{un}_{\mathcal{G}} = \text{un}_{\tilde{\mathcal{G}}}$  by (ii) or  $\alpha \notin \text{un}_{\mathcal{G}}$ . In the latter case, by construction of  $\tilde{\mathcal{G}}$  there is a directed path  $\alpha \rightarrow \nu_n \rightarrow \dots \rightarrow \nu_0$  in  $\tilde{\mathcal{G}}$ , and every vertex on the path is in  $V \setminus \text{un}_{\mathcal{G}}$ . Hence  $\alpha \in \text{ant}_{\mathcal{G}}(\nu_n)$ , and  $\nu_i \in \text{ant}_{\mathcal{G}}(\nu_{i-1})$  ( $i = 1, \dots, n$ ), so  $\alpha \in \text{ant}_{\mathcal{G}}(\nu_0)$ .

(iv) If  $\beta \rightarrow \alpha$  in  $\tilde{\mathcal{G}}$  then, by the construction of  $\tilde{\mathcal{G}}$ ,  $\alpha \notin \text{un}_{\mathcal{G}}$  and  $\beta \in \text{ant}_{\mathcal{G}}(\alpha) \cup \text{un}_{\mathcal{G}}$ . Hence, by Lemma 3.8(ii),  $\alpha \notin \text{ant}_{\mathcal{G}}(\beta)$  and thus  $\alpha \notin \text{ant}_{\tilde{\mathcal{G}}}(\beta)$ , again by (iii). Similarly, if  $\beta \leftrightarrow \alpha$  in  $\tilde{\mathcal{G}}$  then by construction,  $\alpha \notin \text{un}_{\mathcal{G}} \cup \text{ant}_{\mathcal{G}}(\beta)$ , hence by (iii),  $\alpha \notin \text{ant}_{\tilde{\mathcal{G}}}(\beta)$ . Thus  $\alpha \notin \text{ant}_{\tilde{\mathcal{G}}}(\text{pa}_{\tilde{\mathcal{G}}}(\alpha) \cup \text{sp}_{\tilde{\mathcal{G}}}(\alpha))$ , so (i) in the definition of an ancestral graph holds. By the construction of  $\tilde{\mathcal{G}}$ , if  $\text{ne}_{\tilde{\mathcal{G}}}(\alpha) \neq \emptyset$  then  $\alpha \in \text{un}_{\mathcal{G}}$ , and thus, again by construction,  $\text{sp}_{\tilde{\mathcal{G}}}(\alpha) \cup \text{pa}_{\tilde{\mathcal{G}}}(\alpha) = \emptyset$ , hence (ii) in the definition holds as required.  $\square$

**Theorem 5.6** *If  $\mathcal{G}$  is a maximal ancestral graph with  $r$  pairs of vertices that are not adjacent, and  $\mathcal{G}^*$  is any complete supergraph of  $\mathcal{G}$  then there exists a sequence of maximal ancestral graphs*

$$\mathcal{G}^* \equiv \mathcal{G}_0, \dots, \mathcal{G}_r \equiv \mathcal{G}$$

where  $\mathcal{G}_{i+1}$  is a subgraph of  $\mathcal{G}_i$  containing one less edge  $\epsilon_i$  than  $\mathcal{G}_i$ , and  $\text{un}_{\mathcal{G}_{i+1}} = \text{un}_{\mathcal{G}_i}$ .

The sequence of edges removed,  $\langle \epsilon_0, \dots, \epsilon_{r-1} \rangle$ , is such that no undirected edge is removed after a directed edge and no directed edge is removed after a bi-directed edge.

Two examples of this Theorem are shown in Figure 14. (The existence of at least one complete ancestral supergraph  $\mathcal{G}^*$  of  $\mathcal{G}$  is guaranteed by the previous Lemma.)

**Proof:** Let  $\tilde{E}$  be the set of edges that are in  $\mathcal{G}_0 \equiv \mathcal{G}^*$  but not  $\mathcal{G}$ . Place an ordering  $\prec$  on  $\tilde{E}$  as follows:

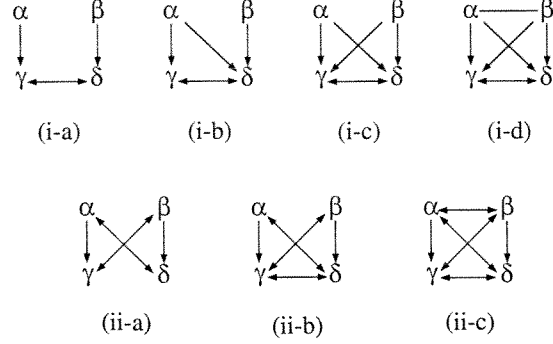


Figure 14: Two simple examples of the extension described in Theorem 5.6. In (ii) if the  $\alpha \leftrightarrow \beta$  edge were added prior to the  $\gamma \leftrightarrow \delta$  edge the resulting graph would not be maximal.

- (i) if  $\alpha - \beta, \gamma \rightarrow \delta \in \tilde{E}$  then  $\alpha - \beta \prec \gamma \rightarrow \delta$ ;
- (ii) if  $\alpha \rightarrow \beta, \gamma \leftrightarrow \delta \in \tilde{E}$  then  $\alpha \rightarrow \beta \prec \gamma \leftrightarrow \delta$ ;
- (iii) if  $\alpha \leftrightarrow \beta, \gamma \leftrightarrow \delta \in \tilde{E}$  and  $\alpha, \beta \in \text{an}_{\mathcal{G}}(\{\gamma, \delta\})$  then  $\alpha \leftrightarrow \beta \prec \gamma \leftrightarrow \delta$ ;

The ordering on bi-directed edges is well-defined by Lemma 3.11. Now let  $\mathcal{G}_i$  be the graph formed by removing the first  $i$  edges in  $\tilde{E}$  under the ordering  $\prec$ . Since  $\mathcal{G}_0$  is ancestral, it follows from Proposition 3.5 that  $\mathcal{G}_i$  is too. Since  $\mathcal{G}_0$  is complete, it is trivially maximal.

Suppose for a contradiction that  $\mathcal{G}_i$  is maximal, but  $\mathcal{G}_{i+1}$  is not. Let the endpoints of  $\epsilon_i$  be  $\alpha$  and  $\beta$ . Since, by hypothesis,  $\mathcal{G}_i$  is maximal, for any pair of vertices  $\gamma, \delta$  that are not adjacent in  $\mathcal{G}_i$ , for some set  $Z$ ,  $(\gamma, \delta \notin Z)$ ,  $\langle \gamma, \delta \mid Z \rangle \in \mathfrak{I}_m(\mathcal{G}_i) \subseteq \mathfrak{I}_m(\mathcal{G}_{i+1})$  (by Proposition 3.12). Since  $\alpha, \beta$  are the only vertices that are not adjacent in  $\mathcal{G}_{i+1}$ , but are adjacent in  $\mathcal{G}_i$ , it follows by Corollaries 4.3 and 4.4 that there is a primitive inducing path  $\pi$  between  $\alpha$  and  $\beta$  in  $\mathcal{G}_{i+1}$  and hence also in  $\mathcal{G}_i$ .

By Corollary 4.6 it then follows that  $\epsilon_i = \alpha \leftrightarrow \beta$  in  $\mathcal{G}_i$ . Since all directed edges in  $\tilde{E}$  occur prior to  $\epsilon_i$ ,  $\text{an}_{\mathcal{G}}(\nu) = \text{an}_{\mathcal{G}_{i+1}}(\nu)$  for all  $\nu \in V$ . By Lemma 4.5 every edge on  $\pi$  is bi-directed and every vertex on the path is in  $\text{an}_{\mathcal{G}_{i+1}}(\{\alpha, \beta\}) = \text{an}_{\mathcal{G}}(\{\alpha, \beta\})$ . It then follows that  $\pi$  exists in  $\mathcal{G}$  since, if any edge on  $\pi$  were in  $\tilde{E}$ , it would occur prior to  $\epsilon_i$ . But in this case, since  $\mathcal{G}$  is maximal,  $\epsilon_i$  is present in  $\mathcal{G}$ , which is a contradiction.

Finally, by Proposition 3.6,  $\text{un}_{\mathcal{G}_i} \subseteq \text{un}_{\mathcal{G}_{i+1}}$ , as  $\mathcal{G}_{i+1}$  is a subgraph of  $\mathcal{G}_i$ . By Lemma 5.5(ii),  $\text{un}_{\mathcal{G}_r} \equiv \text{un}_{\mathcal{G}} = \text{un}_{\tilde{\mathcal{G}}} \equiv \text{un}_{\mathcal{G}_0}$ , hence  $\text{un}_{\mathcal{G}_i} = \text{un}_{\mathcal{G}_{i+1}}$ .  $\square$

Note that the proof shows that between  $\mathcal{G}$  and *any* complete supergraph  $\mathcal{G}_0$  of  $\mathcal{G}$  there will exist a sequence of maximal graphs, each differing from the next by a single edge.

## 6 Canonical Directed Acyclic Graphs

In this section we show that for every maximal ancestral graph  $\mathcal{G}$  there exists a DAG  $\mathcal{D}(\mathcal{G})$  and sets  $S, L$  such that  $\mathcal{D}(\mathcal{G})[L^S] = \mathcal{G}$ . This result is important because it shows that every independence model represented by an ancestral graph corresponds to some DAG model under marginalizing and conditioning.

### 6.1 The Canonical DAG $\mathcal{D}(\mathcal{G})$ Associated with $\mathcal{G}$

If  $\mathcal{G}$  is an ancestral graph with vertex set  $V$ , then we define the *canonical DAG*,  $\mathcal{D}(\mathcal{G})$  associated with  $\mathcal{G}$  as follows:

- (i) let  $S_{\mathcal{D}(\mathcal{G})} = \{\sigma_{\alpha\beta} \mid \alpha - \beta \text{ in } \mathcal{G}\}$
- (ii) let  $L_{\mathcal{D}(\mathcal{G})} = \{\lambda_{\alpha\beta} \mid \alpha \leftrightarrow \beta \text{ in } \mathcal{G}\}$
- (iii) DAG  $\mathcal{D}(\mathcal{G})$  has vertex set  $V \cup L_{\mathcal{D}(\mathcal{G})} \cup S_{\mathcal{D}(\mathcal{G})}$  and edge set defined as follows:

$$\text{If } \left\{ \begin{array}{l} \alpha \rightarrow \beta \\ \alpha \leftrightarrow \beta \\ \alpha - \beta \end{array} \right\} \text{ in } \mathcal{G} \text{ then } \left\{ \begin{array}{l} \alpha \rightarrow \beta \\ \alpha \leftarrow \lambda_{\alpha\beta} \rightarrow \beta \\ \alpha \rightarrow \sigma_{\alpha\beta} \leftarrow \beta \end{array} \right\} \text{ in } \mathcal{D}(\mathcal{G}).$$

Figure 15 shows an ancestral graph and the associated canonical DAG.

Wermuth et al. (1994) introduced the idea of transforming a graph into a DAG in this way by introducing additional ‘synthetic’ variables, as a method of interpreting particular dependence models. (See also Verma and Pearl, 1990.)

A *minipath* is a path in  $\mathcal{D}(\mathcal{G})$  containing one or two edges, with endpoints in  $V$ , but no other vertices in  $V$ . The construction of  $\mathcal{D}(\mathcal{G})$  sets up a one

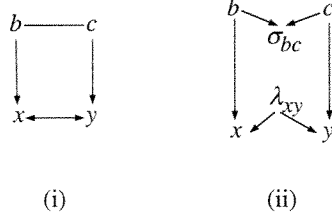


Figure 15: (i) An ancestral graph; (ii) the associated canonical DAG.

to one correspondence between edges in  $\mathcal{G}$ , and minipaths in  $\mathcal{D}(\mathcal{G})$ . If  $\alpha$  and  $\beta$  are adjacent in  $\mathcal{G}$  then denote the corresponding minipath in  $\mathcal{D}(\mathcal{G})$ ,  $\delta_{\alpha\beta}$ . Conversely if  $\delta$  is a minipath in  $\mathcal{D}(\mathcal{G})$ , then let  $\delta^{\mathcal{G}}$  denote the corresponding edge in  $\mathcal{G}$ .

Observe that if  $\delta_{\alpha\beta}$  and  $\delta_{\phi\psi}$  are minipaths corresponding to two different adjacencies in  $\mathcal{G}$ , then no non-endpoint vertices are common to these paths.

Given a path  $\mu$  in  $\mathcal{D}(\mathcal{G})$ , with endpoints in  $V$ , the path may be decomposed into a sequence of minipaths  $\langle \delta_{\alpha_1\alpha_2}, \dots, \delta_{\alpha_{n-1}\alpha_n} \rangle$ , from which we may construct a path  $\langle \alpha_1, \dots, \alpha_n \rangle$  in  $\mathcal{G}$  by replacing each mini-path by the corresponding edge. We will denote this path by  $\mu^{\mathcal{G}}$ . Note that since  $\mathcal{D}(\mathcal{G})$  is a DAG,  $\text{an}_{\mathcal{D}(\mathcal{G})}(\cdot) = \text{ant}_{\mathcal{D}(\mathcal{G})}(\cdot)$ , and by definition a path  $\mu$  is m-connecting if and only if it is d-connecting. Since it helps to make clear that we are referring to a path in a DAG, we will only use the term ‘d-connecting’ when referring to a path which is m-connecting (and d-connecting) in  $\mathcal{D}(\mathcal{G})$ .

### 6.1.1 Graphical properties of $\mathcal{D}(\mathcal{G})$

**Lemma 6.1** *Let  $\mathcal{G}$  be an ancestral graph with vertex set  $V$ .*

- (i) *If  $\beta \in V$  then  $\text{an}_{\mathcal{D}(\mathcal{G})}(\beta) \cap V = \text{an}_{\mathcal{G}}(\beta)$ .*
- (ii)  *$\text{an}_{\mathcal{D}(\mathcal{G})}(S_{\mathcal{D}(\mathcal{G})}) = \text{pa}_{\mathcal{D}(\mathcal{G})}(S_{\mathcal{D}(\mathcal{G})}) \cup S_{\mathcal{D}(\mathcal{G})}$ , so  $\text{an}_{\mathcal{D}(\mathcal{G})}(S_{\mathcal{D}(\mathcal{G})}) \subseteq S_{\mathcal{D}(\mathcal{G})} \cup \text{un}_{\mathcal{G}}$ .*
- (iii)  *$\text{an}_{\mathcal{D}(\mathcal{G})}(S_{\mathcal{D}(\mathcal{G})}) \cap L_{\mathcal{D}(\mathcal{G})} = \emptyset$ .*

**Proof:** (i) If  $\alpha, \beta \in V$  and  $\alpha \in \text{an}_{\mathcal{D}(\mathcal{G})}(\beta)$  then there is a directed path  $\delta$  from  $\alpha$  to  $\beta$  in  $\mathcal{D}(\mathcal{G})$ . Every non-endpoint vertex on  $\delta$  has at least one parent and at least one child in  $\mathcal{D}(\mathcal{G})$ , hence every vertex on  $\delta$  is in  $V$  (since  $\text{ch}_{\mathcal{D}(\mathcal{G})}(S_{\mathcal{D}(\mathcal{G})}) = \emptyset = \text{pa}_{\mathcal{D}(\mathcal{G})}(L_{\mathcal{D}(\mathcal{G})})$ ). It then follows from the construction of

$\mathcal{D}(\mathcal{G})$  that  $\delta$  exists in  $\mathcal{G}$ , so  $\alpha \in \text{an}_{\mathcal{G}}(\beta)$ . It also follows from the construction of  $\mathcal{D}(\mathcal{G})$  that any directed path in  $\mathcal{G}$  exists in  $\mathcal{D}(\mathcal{G})$ .

(ii) By construction,  $\text{pa}_{\mathcal{D}(\mathcal{G})}(\sigma_{\alpha\beta}) = \{\alpha, \beta\} \subseteq \text{un}_{\mathcal{G}}$  (by Lemma 3.7). But again, by construction,  $\text{pa}_{\mathcal{D}(\mathcal{G})}(\text{un}_{\mathcal{G}}) = \emptyset$ . Hence  $\text{an}_{\mathcal{D}(\mathcal{G})}(\sigma_{\alpha\beta}) = \{\alpha, \beta, \sigma_{\alpha\beta}\} \subseteq \text{un}_{\mathcal{G}} \cup \{\sigma_{\alpha\beta}\}$ , so  $\text{an}_{\mathcal{D}(\mathcal{G})}(S_{\mathcal{D}(\mathcal{G})}) \subseteq \text{un}_{\mathcal{G}} \cup S_{\mathcal{D}(\mathcal{G})}$ .

(iii) This follows from the previous property:

$$\text{an}_{\mathcal{D}(\mathcal{G})}(S_{\mathcal{D}(\mathcal{G})}) \cap L_{\mathcal{D}(\mathcal{G})} \subseteq (\text{un}_{\mathcal{G}} \cup S_{\mathcal{D}(\mathcal{G})}) \cap L_{\mathcal{D}(\mathcal{G})} \subseteq (V \cup S_{\mathcal{D}(\mathcal{G})}) \cap L_{\mathcal{D}(\mathcal{G})} = \emptyset$$

□

Note that  $\text{ant}_{\mathcal{G}}(\beta) \neq \text{ant}_{\mathcal{D}(\mathcal{G})}(\beta)$  for  $\beta \in V$ , because an undirected edge  $\alpha - \beta$  in  $\mathcal{G}$  is replaced by  $\alpha \rightarrow \lambda_{\alpha\beta} \leftarrow \beta$  in  $\mathcal{D}(\mathcal{G})$ .

**Lemma 6.2**  $\mathcal{G}$  is a subgraph of  $\mathcal{D}(\mathcal{G})_{L_{\mathcal{D}(\mathcal{G})}}^{S_{\mathcal{D}(\mathcal{G})}}$ .

**Proof:** First recall that  $\text{an}_{\mathcal{D}(\mathcal{G})}(\cdot) = \text{ant}_{\mathcal{D}(\mathcal{G})}(\cdot)$  since  $\mathcal{D}(\mathcal{G})$  is a DAG. We now consider each of the edges occurring in  $\mathcal{G}$ :

(i) If  $\alpha - \beta$  in  $\mathcal{G}$  then  $\alpha \rightarrow \sigma_{\alpha\beta} \leftarrow \beta$  in  $\mathcal{D}(\mathcal{G})$ , so  $\alpha, \beta \in \text{an}_{\mathcal{D}(\mathcal{G})}(S_{\mathcal{D}(\mathcal{G})})$ . It then follows that  $\alpha - \beta$  in  $\mathcal{D}(\mathcal{G})_{L_{\mathcal{D}(\mathcal{G})}}^{S_{\mathcal{D}(\mathcal{G})}}$ .

(ii) If  $\alpha \rightarrow \beta$  in  $\mathcal{G}$  then  $\alpha \rightarrow \beta$  in  $\mathcal{D}(\mathcal{G})$ , so  $\alpha \in \text{ant}_{\mathcal{D}(\mathcal{G})}(\beta)$ . By Lemma 6.1(i),  $\beta \notin \text{ant}_{\mathcal{D}(\mathcal{G})}(\alpha)$ , and since further,  $\beta \notin S_{\mathcal{D}(\mathcal{G})} \cup \text{un}_{\mathcal{G}}$ , by Lemma 6.1(ii),  $\beta \notin \text{ant}_{\mathcal{D}(\mathcal{G})}(S_{\mathcal{D}(\mathcal{G})})$ . It then follows from the definition of the transformation that  $\alpha \rightarrow \beta$  in  $\mathcal{D}(\mathcal{G})_{L_{\mathcal{D}(\mathcal{G})}}^{S_{\mathcal{D}(\mathcal{G})}}$ .

(iii) Likewise, if  $\alpha \leftrightarrow \beta$  in  $\mathcal{G}$  then  $\alpha \leftarrow \lambda_{\alpha\beta} \rightarrow \beta$  in  $\mathcal{D}(\mathcal{G})$ . By Lemma 6.1(i) and (ii), it follows as in case (ii) that  $\beta \notin \text{ant}_{\mathcal{D}(\mathcal{G})}(\{\alpha\} \cup S_{\mathcal{D}(\mathcal{G})})$ , and by symmetry,  $\alpha \notin \text{ant}_{\mathcal{D}(\mathcal{G})}(\{\beta\} \cup S_{\mathcal{D}(\mathcal{G})})$ . Hence  $\alpha \leftrightarrow \beta$  in  $\mathcal{D}(\mathcal{G})_{L_{\mathcal{D}(\mathcal{G})}}^{S_{\mathcal{D}(\mathcal{G})}}$ .

□



## 6.2 The independence model $\mathfrak{I}_m(\mathcal{D}(\mathcal{G})_{[L_{\mathcal{D}(\mathcal{G})}^{S_{\mathcal{D}(\mathcal{G})}}]})$

**Theorem 6.3** *If  $\mathcal{G}$  is an ancestral graph then*

$$\mathfrak{I}_m(\mathcal{G}) = \mathfrak{I}_m(\mathcal{D}(\mathcal{G}))_{[L_{\mathcal{D}(\mathcal{G})}^{S_{\mathcal{D}(\mathcal{G})}}]} = \mathfrak{I}_m(\mathcal{D}(\mathcal{G}))_{[L_{\mathcal{D}(\mathcal{G})}^{S_{\mathcal{D}(\mathcal{G})}}]}.$$

It follows from this result that the global Markov property for ancestral graphs may be reduced to that for DAGs:  $X$  is m-separated from  $Y$  given  $Z$  in  $\mathcal{G}$  if and only if  $X$  is d-separated from  $Y$  given  $Z \cup S_{\mathcal{D}(\mathcal{G})}$ . (However, see Section 8.6 for related comments concerning parametrization.)

It also follows from this result that the class of independence models associated with ancestral graphs is the smallest class that contains the DAG independence models and is closed under marginalizing and conditioning.

We break the proof into three pieces:

**Proof:**  $\mathfrak{I}_m(\mathcal{D}(\mathcal{G}))_{[L_{\mathcal{D}(\mathcal{G})}^{S_{\mathcal{D}(\mathcal{G})}}]} = \mathfrak{I}_m(\mathcal{D}(\mathcal{G}))_{[L_{\mathcal{D}(\mathcal{G})}^{S_{\mathcal{D}(\mathcal{G})}}]}$  by Theorem 4.18.  $\square$

**Proof:**  $\mathfrak{I}_m(\mathcal{G}) \subseteq \mathfrak{I}_m(\mathcal{D}(\mathcal{G}))_{[L_{\mathcal{D}(\mathcal{G})}^{S_{\mathcal{D}(\mathcal{G})}}]}$

Suppose  $\mathcal{G}$  has vertex set  $V$ , containing vertices  $\alpha, \beta$ , and set  $Z$  ( $\alpha, \beta \notin Z$ ). It is sufficient to prove that if there is a path  $\mu$  which d-connects  $\alpha$  and  $\beta$  given  $Z \cup S_{\mathcal{D}(\mathcal{G})}$  in  $\mathcal{D}(\mathcal{G})$  then  $\mu^{\mathcal{G}}$  m-connects  $\alpha$  and  $\beta$  given  $Z$  in  $\mathcal{G}$ .

Suppose that  $\gamma$  is a collider on  $\mu^{\mathcal{G}}$ . In this case  $\gamma$  is a collider on  $\mu$  since the corresponding minipaths collide at  $\gamma$  in  $\mathcal{D}(\mathcal{G})$ . Since  $\mu$  is d-connecting given  $Z \cup S_{\mathcal{D}(\mathcal{G})}$  and  $\gamma \in V$ ,

$$\gamma \in \left( \text{an}_{\mathcal{D}(\mathcal{G})}(Z \cup S_{\mathcal{D}(\mathcal{G})}) \right) \cap V = \left( \text{an}_{\mathcal{D}(\mathcal{G})}(Z) \cap V \right) \cup \left( \text{an}_{\mathcal{D}(\mathcal{G})}(S_{\mathcal{D}(\mathcal{G})}) \cap V \right),$$

by Proposition 2.1. But  $\gamma \notin \text{un}_{\mathcal{G}}$ , so by Lemma 6.1 (ii),  $\gamma \notin \text{an}_{\mathcal{D}(\mathcal{G})}(S_{\mathcal{D}(\mathcal{G})})$ . Hence  $\gamma \in (\text{an}_{\mathcal{D}(\mathcal{G})}(Z) \cap V) = \text{an}_{\mathcal{G}}(Z)$ , the equality following from Lemma 6.1(i).

If  $\gamma$  is a non-collider on  $\mu^{\mathcal{G}}$  then  $\gamma$  is a non-collider on  $\mu$ , so  $\gamma \notin Z \cup S_{\mathcal{D}(\mathcal{G})}$ , thus  $\gamma \notin Z$  as required.  $\square$

**Proof:**  $\mathfrak{I}_m(\mathcal{D}(\mathcal{G}))_{[L_{\mathcal{D}(\mathcal{G})}^{S_{\mathcal{D}(\mathcal{G})}}]} \subseteq \mathfrak{I}_m(\mathcal{G})$

By Lemma 6.2  $\mathcal{G}$  is a subgraph of  $\mathcal{D}(\mathcal{G})_{[L_{\mathcal{D}(\mathcal{G})}^{S_{\mathcal{D}(\mathcal{G})}}]}$ , the result then follows by Proposition 3.12.  $\square$

### 6.2.1 If $\mathcal{G}$ is maximal then $\mathcal{D}(\mathcal{G})_{L_{\mathcal{D}(\mathcal{G})}}^{S_{\mathcal{D}(\mathcal{G})}} = \mathcal{G}$

We now prove the result mentioned at the start of this section:

**Theorem 6.4** *If  $\mathcal{G}$  is a maximal ancestral graph then*

$$\mathcal{D}(\mathcal{G})_{L_{\mathcal{D}(\mathcal{G})}}^{S_{\mathcal{D}(\mathcal{G})}} = \mathcal{G}.$$

**Proof:** By Lemma 6.2  $\mathcal{G}$  is a subgraph of  $\mathcal{D}(\mathcal{G})_{L_{\mathcal{D}(\mathcal{G})}}^{S_{\mathcal{D}(\mathcal{G})}}$ , while by Theorem 6.3 these graphs correspond to the same independence model. It then follows from the maximality of  $\mathcal{G}$  that  $\mathcal{D}(\mathcal{G})_{L_{\mathcal{D}(\mathcal{G})}}^{S_{\mathcal{D}(\mathcal{G})}} = \mathcal{G}$ .  $\square$

## 7 Probability Distributions

In this section we relate the operations of marginalizing and conditioning that have been defined for independence models and graphs to probability distributions.

### 7.1 Marginalizing and Conditioning Distributions

For a graph  $\mathcal{G}$  with vertex set  $V$  we consider collections of random variables  $(X_\nu)_{\nu \in V}$  taking values in probability spaces  $(\mathfrak{X}_\nu)_{\nu \in V}$ . In all the examples we consider, the probability spaces are either real finite-dimensional vector spaces or finite discrete sets. For  $A \subseteq V$  we let  $\mathfrak{X}_A \equiv \times_{\nu \in A} (\mathfrak{X}_\nu)$ ,  $\mathfrak{X} \equiv \mathfrak{X}_V$  and  $X_A \equiv (X_\nu)_{\nu \in A}$ .

If  $P$  is a probability measure on  $\mathfrak{X}_V$  then as usual we define the distribution after marginalizing over  $X_L$ , here denoted  $P_{[X_L]}$  or  $P_{X_{V \setminus L}}$ , to be a probability measure on  $\mathfrak{X}_{V \setminus L}$ , such that

$$P_{[X_L]}(E) \equiv P_{X_{V \setminus L}}(E) = P\left(\langle X_{V \setminus L}, X_L \rangle \in E \times \mathfrak{X}_L\right)$$

We will assume the existence of a regular conditional probability measure, denoted  $P^{[X_S=x_S]}(\cdot)$  or  $P(\cdot | X_S = x_S)$ , for all  $x_S \in \mathfrak{X}_S$  so that

$$\int_F P^{[X_S=x_S]}(E) dP_{X_S}(x_S) = P\left(\langle X_{V \setminus S}, X_S \rangle \in E \times F\right).$$

This defines  $P^{[X_S=x_S]}(\cdot)$  up to almost sure equivalence under  $P_{X_S}$ . Likewise we define

$$P_{[X_L]}^{[X_S=x_S]}(\cdot) \equiv (P^{[X_S=x_S]}_{[X_L]}(\cdot)).$$

## 7.2 The Set of Distributions Obeying an Independence Model $(\mathcal{P}(\mathfrak{I}))$

We define conditional independence under  $P$  as follows:

$$A \perp\!\!\!\perp B \mid C [P] \iff P_{X_{V \setminus (A \cup C)}}^{X_C=x_C}(\cdot) = P_{X_{V \setminus (A \cup B \cup C)}}^{X_B=x_B, X_C=x_C}(\cdot) \quad (P_{X_{B \cup C}} \text{ a.e.})$$

where we have used the usual shorthand notation:  $A$  denotes both a vertex set and the random variable  $X_A$ .

For an independence model  $\mathfrak{I}$  over  $V$  let  $\mathcal{P}(\mathfrak{I})$  be the set of distributions  $P$  on  $\mathfrak{X}$  such that for arbitrary disjoint sets  $A, B, Z$ , ( $Z$  may be empty)

$$\text{if } \langle A, B \mid Z \rangle \in \mathfrak{I} \text{ then } A \perp\!\!\!\perp B \mid Z [P].$$

Note that if  $P \in \mathcal{P}(\mathfrak{I})$  then there may be independence relations that are not in  $\mathfrak{I}$  that also hold in  $P$ .

A distribution  $P$  is said to be *faithful* or *Markov perfect* with respect to an independence model  $\mathfrak{I}$  if

$$\langle A, B \mid Z \rangle \in \mathfrak{I} \text{ if and only if } A \perp\!\!\!\perp B \mid Z [P].$$

An independence model  $\mathfrak{I}$  is said to be *probabilistic* if there is a distribution  $P$  that is faithful to  $\mathfrak{I}$ .

## 7.3 Relating $\mathcal{P}(\mathfrak{I}_m(\mathcal{G}))$ and $\mathcal{P}(\mathfrak{I}_m(\mathcal{G}_L^S))$

**Theorem 7.1** *Let  $\mathfrak{I}$  be an independence model over  $V$  with  $S \cup L \subset V$ . If  $P \in \mathcal{P}(\mathfrak{I})$  then*

$$P_{X_L}^{X_S=x_S} \in \mathcal{P}(\mathfrak{I}_L^S) \quad (P_{X_S} \text{ a.e.}).$$

**Proof:** Suppose  $\langle X, Y \mid Z \rangle \in \mathfrak{I}_L^S$ . It follows that  $\langle X, Y \mid Z \cup S \rangle \in \mathfrak{I}$  and  $(X \cup Y \cup Z) \subseteq V \setminus (S \cup L)$ . Hence, if  $P \in \mathcal{P}(\mathfrak{I})$  and  $\langle X, Y \mid Z \rangle \in \mathfrak{I}_L^S$  then

$$X \perp\!\!\!\perp Y \mid Z \cup S [P],$$

hence

$$X \perp\!\!\!\perp Y \mid Z \left[ P_{X_L}^{X_S=x_S} \right] \quad (P_{X_S} \text{ a.e.}).$$

(The last step follows from the assumption that regular conditional probability measures exist. See Koster (1999a) Appendix A & B.) Since there are finitely many triples  $\langle X, Y \mid Z \rangle \in \mathfrak{I}_L^S$ , it follows that

$$P_{x_L}^{x_S=x_S} \in \mathcal{P}(\mathfrak{I}_L^S) \quad (P_{X_S} \text{ a.e.}),$$

as required.  $\square$

Two corollaries follow from this result:

**Corollary 7.2** *If  $\mathcal{G}$  is an ancestral graph and  $P \in \mathcal{P}(\mathfrak{I}_m(\mathcal{G}))$  then*

$$P_{x_L}^{x_S=x_S} \in \mathcal{P}(\mathfrak{I}_m(\mathcal{G})_L^S) = \mathcal{P}(\mathfrak{I}_m(\mathcal{G}_L^S)) \quad (P_{X_S} \text{ a.e.}).$$

**Proof:** This follows directly from Theorem 7.1 and Theorem 4.18.  $\square$

**Corollary 7.3** *If  $N$  is a normal distribution, faithful to an independence model  $\mathfrak{I}$  over vertex set  $V$  then  $N_{x_L}^{x_S=x_S}$  is faithful to  $\mathfrak{I}_L^S$ .*

**Proof:** Since  $N \in \mathcal{P}(\mathfrak{I})$ , by normality and Theorem 7.1,  $N_{x_L}^{x_S=x_S} \in \mathcal{P}(\mathfrak{I}_L^S)$ . Now suppose  $\langle X, Y \mid Z \rangle \notin \mathfrak{I}_L^S$  where  $X \dot{\cup} Y \dot{\cup} Z \subseteq V \setminus (S \cup L)$ . Hence  $\langle X, Y \mid Z \cup S \rangle \notin \mathfrak{I}$ . Since  $N$  is faithful to  $\mathfrak{I}$ ,

$$X \not\perp\!\!\!\perp Y \mid Z \cup S \quad [N] \quad \text{which implies} \quad X \not\perp\!\!\!\perp Y \mid Z \quad [N_{x_L}^{x_S=x_S}]$$

for any  $x_S \in \mathbb{R}^{|S|}$ , by standard properties of the normal distribution.  $\square$

Note that the analogous result is not true for the multinomial distribution as context-specific (or asymmetric) independence relations may be present.

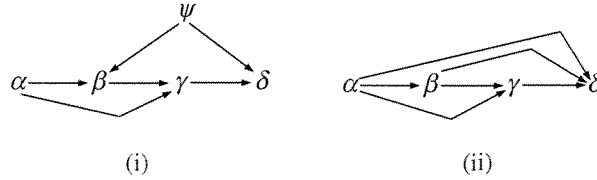


Figure 16: (i) An ancestral graph  $\mathcal{G}$ ; (ii) the graph  $\mathcal{G}_{\{\psi\}}^{\emptyset}$ . (See Section 7.3.1.)

### 7.3.1 A non-independence restriction

The following example due to Verma and Pearl (1991), Robins (1997) shows that there are distributions  $Q \in \mathcal{P}(\mathcal{I}_m(\mathcal{G})_L^S)$  for which there is no distribution  $P \in \mathcal{P}(\mathcal{I}_m(\mathcal{G}))$  such that  $Q = P_L^S$ . In other words, a set of distributions defined via a set of independence relations may impose constraints on a given margin that are not independence relations.

Consider the graph  $\mathcal{G}$  in Figure 16(i). Marginalizing over  $\psi$  produces the complete graph  $\mathcal{G}_{\{\psi\}}^\emptyset$  shown in Figure 16(ii), so  $\mathcal{P}(\mathcal{I}_m(\mathcal{G}_{\{\psi\}}^\emptyset))$  is the saturated model containing every distribution over  $\{\alpha, \beta, \gamma, \delta\}$ . However, if  $P \in \mathcal{P}(\mathcal{I}_m(\mathcal{G}))$  then, almost surely under  $P(X_\alpha, X_\gamma)$ ,

$$\begin{aligned}
& \int_{\mathfrak{X}_\beta} P(X_\delta \mid x_\alpha, x_\beta, x_\gamma) dP(x_\beta \mid x_\alpha) \\
&= \int_{\mathfrak{X}_\beta} \int_{\mathfrak{X}_\psi} P(X_\delta \mid x_\alpha, x_\beta, x_\gamma, x_\psi) dP(x_\psi \mid x_\alpha, x_\beta, x_\gamma) dP(x_\beta \mid x_\alpha) \\
&= \int_{\mathfrak{X}_\beta} \int_{\mathfrak{X}_\psi} P(X_\delta \mid x_\alpha, x_\beta, x_\gamma, x_\psi) dP(x_\psi \mid x_\alpha, x_\beta) dP(x_\beta \mid x_\alpha) \\
&\hspace{15em} \text{since } \gamma \perp\!\!\!\perp \psi \mid \{\alpha, \beta\} \\
&= \int_{\mathfrak{X}_\beta \times \mathfrak{X}_\psi} P(X_\delta \mid x_\alpha, x_\beta, x_\gamma, x_\psi) dP(x_\beta, x_\psi \mid x_\alpha) \\
&= \int_{\mathfrak{X}_\beta \times \mathfrak{X}_\psi} P(X_\delta \mid x_\alpha, x_\gamma, x_\psi) dP(x_\beta, x_\psi \mid x_\alpha) \hspace{1em} \text{since} \\
&\hspace{15em} \beta \perp\!\!\!\perp \delta \mid \{\alpha, \gamma, \psi\} \\
&= \int_{\mathfrak{X}_\psi} P(X_\delta \mid x_\alpha, x_\gamma, x_\psi) dP(x_\psi \mid x_\alpha) \\
&= \int_{\mathfrak{X}_\psi} P(X_\delta \mid x_\gamma, x_\psi) dP(x_\psi) \hspace{1em} \text{since } \alpha \perp\!\!\!\perp \psi, \\
&\hspace{15em} \alpha \perp\!\!\!\perp \delta \mid \{\gamma, \psi\}.
\end{aligned}$$

This will not hold in general for an arbitrary distribution since the last expression is not a function of  $x_\alpha$ . However, faithfulness is preserved under marginalization for arbitrary distributions.

## 7.4 Independence Models for Ancestral Graphs Are Probabilistic

The existence of distributions that are faithful to  $\mathfrak{I}_m(\mathcal{G})$  for an ancestral graph  $\mathcal{G}$  follows from the corresponding result for DAGs:

**Theorem 7.4** (Building on results of Geiger, 1990, Geiger and Pearl, 1990, Frydenberg, 1990b, Spirtes et al., 1993 and Meek, 1995b.)

*For an arbitrary DAG,  $\mathcal{D}$ ,  $\mathfrak{I}_m(\mathcal{D})$  is probabilistic, in particular there is a normal distribution that is faithful to  $\mathfrak{I}_m(\mathcal{D})$ .*

**Theorem 7.5** *If  $\mathcal{G}$  is an ancestral graph then  $\mathfrak{I}_m(\mathcal{G})$  is probabilistic, in particular there is a normal distribution which is faithful to  $\mathfrak{I}_m(\mathcal{G})$ .*

**Proof:** By Theorem 6.3 there is a DAG  $\mathcal{D}(\mathcal{G})$  such that

$$\mathfrak{I}_m(\mathcal{G}) = \mathfrak{I}_m(\mathcal{D}(\mathcal{G}))_{[L_{\mathcal{D}(\mathcal{G})}]^{S_{\mathcal{D}(\mathcal{G})}}}$$

By Theorem 7.4 there is a normal distribution  $N$  that is faithful to  $\mathfrak{I}_m(\mathcal{D}(\mathcal{G}))$ . By Corollary 7.3,  $N_{[X_L^{X_S=X_S}]}^{X_S=X_S}$  is faithful to  $\mathfrak{I}_m(\mathcal{D}(\mathcal{G}))_{[L_{\mathcal{D}(\mathcal{G})}]^{S_{\mathcal{D}(\mathcal{G})}}} = \mathfrak{I}_m(\mathcal{D}(\mathcal{G}))_{[L_{\mathcal{D}(\mathcal{G})}]^{S_{\mathcal{D}(\mathcal{G})}}} = \mathfrak{I}_m(\mathcal{G})$ .  $\square$

### 7.4.1 Completeness of the global Markov property

A graphical separation criterion  $C$  is said to be *complete* if for any graph  $\mathcal{G}$  and independence model  $\mathfrak{I}^*$ ,

$$\text{if } \mathfrak{I}_C(\mathcal{G}) \subseteq \mathfrak{I}^* \text{ and } \mathcal{P}(\mathfrak{I}_C(\mathcal{G})) = \mathcal{P}(\mathfrak{I}^*) \text{ then } \mathfrak{I}_C(\mathcal{G}) = \mathfrak{I}^*.$$

In other words, the independence model  $\mathfrak{I}_C(\mathcal{G})$  (see Section 2.1.1) cannot be extended without changing the associated set of distributions  $\mathcal{P}(\mathfrak{I}_C(\mathcal{G}))$ .

**Theorem 7.6** *The global Markov property for ancestral graphs is complete.*

**Proof:** The existence of a distribution that is faithful to  $\mathfrak{I}_m(\mathcal{G})$  is clearly a sufficient condition for completeness.  $\square$

## 8 Gaussian Parametrization

There is a natural parametrization of the set of *all* non-singular normal distributions satisfying the independence relations in  $\mathfrak{I}_m(\mathcal{G})$ . In the following sections we first introduce the parametrization, then define the set of normal distributions satisfying the relations in the independence model, and then prove equivalence.

Let  $N_p(\boldsymbol{\mu}, \Sigma)$  denote a  $p$ -dimensional multivariate normal distribution with mean  $\boldsymbol{\mu}$  and covariance matrix  $\Sigma$ . Likewise let  $\mathcal{N}_p$  be the set of all such distributions, with non-singular covariance matrices.

Throughout this section we find it useful to make the following convention:  $\Sigma_{AA}^{-1} = (\Sigma_{AA})^{-1}$ , where  $\Sigma_{AA}$  is the submatrix of  $\Sigma$  restricted to  $A$ .

### 8.1 Parametrization

A *Gaussian parametrization* of an ancestral graph  $\mathcal{G}$ , with vertex set  $V$  and edge set  $E$  is a pair  $\langle \mu, \Phi \rangle$ , consisting of a *mean function*

$$\mu : V \rightarrow \mathbb{R}$$

which assigns a number to every vertex, together with a *covariance function*

$$\Phi : V \cup E \rightarrow \mathbb{R}$$

which assigns a number to every edge *and* vertex in  $\mathcal{G}$ , subject to the restriction that the matrices  $\Lambda, \Omega$  defined below are positive definite (p.d.):

$$(\Lambda)_{\alpha\beta} = \lambda_{\alpha\beta} = \begin{cases} \Phi(\alpha) & \text{if } \alpha = \beta, \\ \Phi(\alpha - \beta) & \text{if } \alpha - \beta \text{ in } \mathcal{G}, \\ 0 & \text{otherwise;} \end{cases}$$

$$(\Omega)_{\alpha\beta} = \omega_{\alpha\beta} = \begin{cases} \Phi(\alpha) & \text{if } \alpha = \beta, \\ \Phi(\alpha \leftrightarrow \beta) & \text{if } \alpha \leftrightarrow \beta \text{ in } \mathcal{G}, \\ 0 & \text{otherwise.} \end{cases}$$

Let  $\Phi(\mathcal{G})$  be the set of all such parametrizations  $\langle \mu, \Phi \rangle$  for  $\mathcal{G}$ . We further define:

$$(B)_{\alpha\beta} = b_{\alpha\beta} = \begin{cases} 1 & \text{if } \alpha = \beta \\ \Phi(\alpha \leftarrow \beta) & \text{if } \alpha \leftarrow \beta \text{ in } \mathcal{G} \\ 0 & \text{otherwise.} \end{cases}$$

**Proposition 8.1** *If  $\Lambda$  and  $\Omega$  are given by a parametrization of  $\mathcal{G}$  then*

- (i)  $\Lambda, \Omega$  are symmetric;
- (ii) for  $\nu \in \text{un}_{\mathcal{G}}$ ,  $\lambda_{\nu\nu} > 0$  and for  $\nu \in V \setminus \text{un}_{\mathcal{G}}$ ,  $\omega_{\nu\nu} > 0$ .

**Proof:** Both properties follow from the requirement that  $\Lambda, \Omega$  be positive definite.  $\square$

**Proposition 8.2** *Let  $\mathcal{G}$  be an ancestral graph with vertices  $V$ , edges  $E$ . The values taken by  $\Phi(\cdot)$  on the sets,  $\text{un}_{\mathcal{G}} \cup \{\alpha - \beta \in E\}$ ,  $(V \setminus \text{un}_{\mathcal{G}}) \cup \{\alpha \leftrightarrow \beta \in E\}$  and  $\{\alpha \rightarrow \beta \in E\}$  are variation independent as  $\Phi(\cdot)$  varies in  $\Phi(\mathcal{G})$ . Likewise,  $\mu(\cdot)$  and  $\Phi(\cdot)$  are variation independent.*

**Proof:** Follows directly from the definition of a parametrization.  $\square$

**Lemma 8.3** *Let  $\mathcal{G}$  be an ancestral graph with vertex set  $V$ . Further, let  $\prec$  be an arbitrary ordering of  $V$  such that all vertices in  $\text{un}_{\mathcal{G}}$  precede those in  $V \setminus \text{un}_{\mathcal{G}}$ , and  $\alpha \in \text{an}(\beta) \setminus \{\beta\}$  implies  $\alpha \prec \beta$ . Under such an ordering, the matrix  $B$  given by a parametrization of  $\mathcal{G}$  has the form:*

$$B = \begin{pmatrix} I & 0 \\ B_{du} & B_{dd} \end{pmatrix}, \quad \text{and} \quad B^{-1} = \begin{pmatrix} I & 0 \\ -B_{dd}^{-1}B_{du} & B_{dd}^{-1} \end{pmatrix},$$

where  $B_{dd}$  is lower triangular, with diagonal entries equal to 1. Hence  $B$  is lower triangular and non-singular, as is  $B^{-1}$ .

Note that we use  $u, d$  as abbreviations for  $\text{un}_{\mathcal{G}}, V \setminus \text{un}_{\mathcal{G}}$  respectively.

**Proof:** If  $\alpha, \beta \in \text{un}_{\mathcal{G}}$  then since  $\mathcal{G}$  is ancestral,  $\alpha \notin \text{ch}_{\mathcal{G}}(\beta)$  and vice versa. Hence by definition of  $B$ ,  $b_{\alpha\beta} = \delta(\alpha, \beta)$  (where  $\delta$  is Kronecker's delta function). If  $\alpha \in \text{un}_{\mathcal{G}}, \beta \in V \setminus \text{un}_{\mathcal{G}}$  then  $\alpha \notin \text{ch}_{\mathcal{G}}(\beta)$ , since  $\mathcal{G}$  is ancestral, hence  $b_{\alpha\beta} = 0$ . If  $\alpha, \beta \in V \setminus \text{un}_{\mathcal{G}}$ , and  $\alpha = \beta$  then  $b_{\alpha\beta} = 1$  by definition. If  $\alpha \neq \beta$ , and  $b_{\alpha\beta} \neq 0$  then  $\alpha \in \text{ch}_{\mathcal{G}}(\beta)$ , so  $\beta \prec \alpha$ . Finally, since  $\mathcal{G}$  is ancestral,  $\beta \notin \text{ch}_{\mathcal{G}}(\alpha)$ , so  $b_{\beta\alpha} = 0$  as required.  $\square$



### 8.1.1 Definition of the Gaussian model $\mathcal{N}(\mathcal{G})$

A parametrization  $\langle \mu, \Phi \rangle$  of  $\mathcal{G}$  specifies a Gaussian distribution as follows:

$$N_{\mathcal{G}\mu\Phi} = N_{|V|}(\boldsymbol{\mu}, \Sigma_{\mathcal{G}\Phi})$$

where

$$(\boldsymbol{\mu})_{\alpha} = \mu(\alpha), \text{ and } \Sigma_{\mathcal{G}\Phi} = B^{-1} \begin{pmatrix} \Lambda^{-1} & 0 \\ 0 & \Omega \end{pmatrix} B^{-\top}. \quad (1)$$

The *Gaussian model*,  $\mathcal{N}(\mathcal{G})$  associated with  $\mathcal{G}$  is the set of normal distributions obtained from parametrizations of  $\mathcal{G}$ :

$$\mathcal{N}(\mathcal{G}) = \{N_{\mathcal{G}\mu\Phi} \mid \langle \mu, \Phi \rangle \in \Phi(\mathcal{G})\}$$

The mean function  $\mu$  does not play a significant role in what follows.

**Lemma 8.4** *If  $\langle \mu, \Phi \rangle$  is a parametrization of an ancestral graph  $\mathcal{G}$  then*

$$\begin{aligned} \Sigma_{\mathcal{G}\Phi} &= \begin{pmatrix} \Lambda^{-1} & -\Lambda^{-1}B_{du}^{\top}B_{dd}^{-\top} \\ -B_{dd}^{-1}B_{du}\Lambda^{-1} & B_{dd}^{-1}(B_{du}\Lambda^{-1}B_{du}^{\top} + \Omega)B_{dd}^{-\top} \end{pmatrix}, \\ \Sigma_{\mathcal{G}\Phi}^{-1} &= \begin{pmatrix} \Lambda + B_{du}^{\top}\Omega^{-1}B_{du} & B_{du}^{\top}\Omega^{-1}B_{dd} \\ B_{dd}^{\top}\Omega^{-1}B_{du} & B_{dd}^{\top}\Omega^{-1}B_{dd} \end{pmatrix}. \end{aligned}$$

**Proof:** Immediate from the definition of  $\Sigma_{\mathcal{G}\Phi}$  and Lemma 8.3.  $\square$

Note that it follows from the conditions on  $B$ ,  $\Lambda$  and  $\Omega$  that  $\Sigma_{\mathcal{G}\Phi}$  is positive definite. The mean function  $\mu$  does not play a significant role in what follows.

### 8.1.2 Parametrization of a subgraph

**Lemma 8.5** *Let  $\langle \mu, \Phi \rangle$  be a parametrization of an ancestral graph  $\mathcal{G} = (V, E)$ . If  $A \subset V$  such that  $\text{ant}(A) = A$ , and  $\langle \mu_A, \Phi_A \rangle$  is the parametrization of the induced subgraph  $\mathcal{G}_A$ , obtained by restricting  $\mu$  to  $A$  and  $\Phi$  to  $A \cup E^*$ , where  $E^*$  is the set of edges in  $\mathcal{G}_A$ , then*

$$\Lambda_A^{-1} = (\Lambda^{-1})_{A \cap u \ A \cap u}, \quad \Omega_A = (\Omega)_{A \cap d \ A \cap d}, \quad B_A^{-1} = (B^{-1})_{AA},$$

hence

$$\Sigma_{\mathcal{G}_A\Phi_A} = (\Sigma_{\mathcal{G}\Phi})_{AA},$$

where  $\Lambda_A, \Omega_A, B_A$  are the matrices associated with  $\Phi_A$ .

In words, if all vertices that are anterior to a set  $A$  in  $\mathcal{G}$  are contained in  $A$  then the covariance matrix parametrized by the restriction of  $\Phi$  to the induced subgraph  $\mathcal{G}_A$  is just the sub-matrix  $(\Sigma_{\mathcal{G}\Phi})_{AA}$ .

Note the distinction between matrices indexed by two subsets which indicate submatrices in the usual way (e.g.  $\tilde{\Sigma}_{AA}$ ) and matrices indexed by one subset which are obtained from a parametrization of an induced subgraph on this set of vertices (e.g.  $B_A$ ).

**Proof:** For  $\Omega$  there is nothing to prove. Since  $A = \text{ant}(A)$ , no vertex in  $\text{un}_{\mathcal{G}} \cap A$  is adjacent to a vertex in  $\text{un}_{\mathcal{G}} \setminus A$ . Thus

$$\Lambda = \begin{pmatrix} \Lambda_{A \cap u \ A \cap u} & 0 \\ 0 & \Lambda_{u \setminus A \ u \setminus A} \end{pmatrix}$$

so  $(\Lambda^{-1})_{A \cap u \ A \cap u} = \Lambda_A^{-1}$  as required.

Since  $A$  is anterior,

$$B = \begin{pmatrix} B_{AA} & 0 \\ B_{\overline{AA}} & B_{\overline{AA}} \end{pmatrix},$$

where  $\overline{A} = V \setminus A$ . The result then follows by partitioned inversion since  $B_A = (B)_{AA} = B_{AA}$ .  $\square$

If  $\mathcal{G} = (V, E)$  is a subgraph of an ancestral graph  $\mathcal{G}^* = (V, E^*)$ , then there is a natural mapping  $\langle \mu, \Phi \rangle \mapsto \langle \mu^*, \Phi^* \rangle$  from  $\Phi(\mathcal{G})$  to  $\Phi(\mathcal{G}^*)$ , defined by:

$$\mu^*(\cdot) = \mu(\cdot) \quad \Phi^*(x) = \begin{cases} \Phi(x) & \text{if } x \in V \cup E, \\ 0 & \text{if } x \in E^* \setminus E. \end{cases}$$

$\Phi^*$  simply assigns 0 to edges in  $\mathcal{G}^*$  that are not in  $\mathcal{G}$  (both graphs have the same vertex set). It is simple to see that

$$N_{\mathcal{G}\mu\Phi} = N_{\mathcal{G}^*\mu^*\Phi^*}.$$

The next Proposition is an immediate consequence:

**Proposition 8.6** *If  $\mathcal{G} = (V, E)$  is a subgraph of an ancestral graph  $\mathcal{G}^* = (V, E^*)$  then  $\mathcal{N}(\mathcal{G}) \subseteq \mathcal{N}(\mathcal{G}^*)$ .*

### 8.1.3 Interpretation of parameters

**Theorem 8.7** *If  $\mathcal{G} = (V, E)$  is an ancestral graph,  $\langle \mu, \Phi \rangle \in \Phi(\mathcal{G})$ , and  $\tilde{\Sigma} = \Sigma_{\mathcal{G}\Phi}$ , then for all vertices  $\alpha$  for which  $\text{pa}(\alpha) \neq \emptyset$ ,*

$$B_{\{\alpha\}\text{pa}(\alpha)} = -\tilde{\Sigma}_{\{\alpha\}\text{pa}(\alpha)}\tilde{\Sigma}_{\text{pa}(\alpha)\text{pa}(\alpha)}^{-1}; \quad \text{further,} \quad \begin{pmatrix} \Lambda^{-1} & 0 \\ 0 & \Omega \end{pmatrix} = B\tilde{\Sigma}B^\top. \quad (2)$$

Regarding  $\tilde{\Sigma}$  as the covariance matrix for a (normal) random vector  $X_V$ , the Theorem states that  $\Phi(\alpha \leftarrow \nu)$  is  $-1$  times the coefficient of  $X_\nu$  in the regression of  $X_\alpha$  on  $X_{\text{pa}(\alpha)}$ .  $\Omega$  is the covariance matrix of the residuals from this set of regressions.  $\Lambda$  is just the inverse covariance matrix for  $X_{\text{un}\mathcal{G}}$ . Hence if  $\tilde{\Sigma}$  is obtained from some unknown covariance function  $\Phi$  for an ancestral graph  $\mathcal{G}$ , then equation (2) allows us to reconstruct  $\Phi$  from  $\mathcal{G}$  and  $\tilde{\Sigma}$ .

**Proof:** Suppose that  $\tilde{\Sigma} = \Sigma_{\mathcal{G}\Phi}$  for some parametrization  $\langle \mu, \Phi \rangle$ . If every vertex has no parents then  $B$  is the identity matrix and the claim holds trivially.

Suppose that  $\alpha$  is a vertex with  $\text{pa}(\alpha) \neq \emptyset$ , hence by definition  $\alpha \in V \setminus \text{un}\mathcal{G}$ . Let  $A = \text{ant}(\alpha)$ ,  $e = \text{ant}(\alpha) \setminus (\text{pa}(\alpha) \cup \{\alpha\})$ ,  $p = \text{pa}(\alpha)$ . By Lemma 8.5,

$$\tilde{\Sigma}_{AA} = B_A^{-1} \begin{pmatrix} \Lambda_A^{-1} & 0 \\ 0 & \Omega_A \end{pmatrix} B_A^{-\top}. \quad (3)$$

Since  $\mathcal{G}$  is ancestral,  $\text{ne}_{\mathcal{G}}(\alpha) \cap A = \emptyset$ . Thus partitioning  $A$  into  $e, p, \{\alpha\}$ , we obtain

$$B_A = \begin{pmatrix} B_{ee} & 0 & 0 \\ B_{pe} & B_{pp} & 0 \\ 0 & B_{\alpha p} & 1 \end{pmatrix}, \quad \text{and} \quad \Omega_A = \begin{pmatrix} \Omega_{e \cap d, e \cap d} & \Omega_{e \cap d, p \cap d} & 0 \\ \Omega_{p \cap d, e \cap d} & \Omega_{p \cap d, p \cap d} & 0 \\ 0 & 0 & \omega_{\alpha\alpha} \end{pmatrix}.$$

The expression for  $B_{\{\alpha\}\text{pa}(\alpha)} = B_{\alpha p}$  then follows from (3) by routine calculation. The second claim is an immediate consequence of (1).  $\square$

### 8.1.4 Identifiability

**Corollary 8.8** *If  $\mathcal{G}$  is an ancestral graph,  $\Phi_1, \Phi_2$  are two covariance functions for  $\mathcal{G}$  and  $\Sigma_{\mathcal{G}\Phi_1} = \Sigma_{\mathcal{G}\Phi_2}$  then  $\Phi_1(\cdot) = \Phi_2(\cdot)$ . Hence the mapping  $\Phi \mapsto \Sigma_{\mathcal{G}\Phi}$  is one-to-one.*

**Proof:** This follows directly from Theorem 8.7: both  $\Phi_1$  and  $\Phi_2$  satisfy equation (2) and hence are identical.  $\square$

### 8.1.5 $\mathcal{N}(\mathcal{G})$ for a complete ancestral graph is saturated

**Theorem 8.9** *If  $\mathcal{G} = (V, E)$  is a complete ancestral graph then  $\mathcal{N}(\mathcal{G}) = \mathcal{N}_{|V|}$ .*

In words, a complete ancestral graph parametrizes the saturated Gaussian model of dimension  $|V|$ .

**Proof:** Let  $\tilde{\Sigma}$  be an arbitrary p.d. matrix of dimension  $|V|$ . It is sufficient to show that there exists a covariance function  $\Phi$  for  $\mathcal{G}$ , such that  $\tilde{\Sigma} = \Sigma_{\mathcal{G}\Phi}$ . We may apply equation (2) to obtain matrices  $B$ ,  $\Lambda$  and  $\Omega$  from  $\tilde{\Sigma}$ . However, it still remains to show that (a) whenever there is a non-zero off-diagonal entry in  $\Lambda$ ,  $\Omega$  or  $B$ , there is an edge of the appropriate type in  $\mathcal{G}$  to associate with it, and (b)  $\Lambda$  and  $\Omega$  are positive definite.

By Lemma 3.21(ii),  $\mathcal{G}_{\text{un}\mathcal{G}}$  is complete, hence in  $\Lambda$  all off-diagonal entries are permitted to be non-zero.

It follows directly from the construction of  $B$  given by (2) that if  $(B)_{\alpha\beta} \neq 0$  and  $\alpha \neq \beta$  then  $\beta \in \text{pa}(\alpha)$ .

Now suppose,  $\alpha, \beta \in V \setminus \text{un}\mathcal{G}$ , and there is no edge  $\alpha \leftrightarrow \beta$  in  $\mathcal{G}$ . Since  $\mathcal{G}$  is complete, it follows from Lemma 3.21(iii) that either  $\alpha \leftarrow \beta$ , or  $\alpha \rightarrow \beta$ . Without loss of generality suppose the former, and let  $A = \text{ant}(\alpha) = \text{pa}(\alpha) \cup \{\alpha\}$  since  $\mathcal{G}$  is complete. Then

$$\begin{aligned} (B\tilde{\Sigma}B^\top)_{\alpha\beta} &= (B_A\tilde{\Sigma}_{AA}B_A^\top)_{\alpha\beta} \\ &= \left[ \begin{pmatrix} B_{pp} & 0 \\ -\tilde{\Sigma}_{\alpha p}\tilde{\Sigma}_{pp}^{-1} & 1 \end{pmatrix} \begin{pmatrix} \tilde{\Sigma}_{pp} & \tilde{\Sigma}_{p\alpha} \\ \tilde{\Sigma}_{\alpha p} & \tilde{\Sigma}_{\alpha\alpha} \end{pmatrix} \begin{pmatrix} B_{pp}^\top & -\tilde{\Sigma}_{pp}^{-1}\tilde{\Sigma}_{p\alpha} \\ 0 & 1 \end{pmatrix} \right]_{\alpha\beta} \\ &= 0 \end{aligned}$$

as required. The same argument applies in the case where  $\beta \in \text{un}\mathcal{G}$ ,  $\alpha \in V \setminus \text{un}\mathcal{G}$ , and hence  $\alpha \leftarrow \beta$ , thus establishing that  $B\tilde{\Sigma}B^\top$  is block-diagonal with blocks  $\Lambda^{-1}$  and  $\Omega$ . This establishes (a).

Since, by hypothesis,  $\tilde{\Sigma}$  is p.d. and  $B$  is non-singular, by construction, it follows that  $\Lambda$  and  $\Omega$  are also p.d. hence (b) holds. We now have:

$$\Sigma_{\mathcal{G}\Phi} = B^{-1} \begin{pmatrix} \Lambda^{-1} & 0 \\ 0 & \Omega \end{pmatrix} B^{-\top} = B^{-1}B\tilde{\Sigma}B^\top B^{-\top} = \tilde{\Sigma}.$$

□

### 8.1.6 Entries in $\Omega^{-1}$ and $\mathcal{G}_{\leftrightarrow}$

If  $\mathcal{G} = (V, E)$  is an ancestral graph then we define  $\mathcal{G}_{\leftrightarrow}$  to be the induced subgraph with vertex set  $V$ , but including only the bi-directed edges in  $E$ .

**Lemma 8.10** *If  $\alpha, \beta \in V \setminus \text{un}_{\mathcal{G}}$  and  $\alpha$  is not adjacent to  $\beta$  in  $(\mathcal{G}_{\leftrightarrow})^a$  then*

$$(\Omega^{-1})_{\alpha\beta} = 0,$$

*for any  $\Omega$  obtained from a covariance function  $\Phi$  for  $\mathcal{G}$ .*

**Proof:** (Based on the proof of Lemma 3.1.6 in Koster (1999a).)

First recall that  $\alpha$  and  $\beta$  are adjacent in  $(\mathcal{G}_{\leftrightarrow})^a$  if and only if  $\alpha$  and  $\beta$  are collider connected in  $\mathcal{G}_{\leftrightarrow}$ . The proof is by induction on  $|d| = |V \setminus \text{un}_{\mathcal{G}}|$ .

If  $|d| = 2$  then  $(\Omega^{-1})_{\alpha\beta} = -(\Omega)_{\alpha\beta}|\Omega|^{-1} = 0$  as there is no edge  $\alpha \leftrightarrow \beta$  in  $\mathcal{G}$ .

For  $|d| > 2$ , note that by partitioned inversion:

$$(\Omega^{-1})_{\alpha\beta} = -(\omega_{\alpha\beta} - \Omega_{\{\alpha\}c} \Omega_{cc}^{-1} \Omega_{c\{\beta\}}) |\Omega_{\{\alpha,\beta\}.c}|^{-1} \quad (4)$$

$$= -\left(\omega_{\alpha\beta} - \sum_{\gamma, \delta \in c} \omega_{\alpha\gamma} (\Omega_{cc}^{-1})_{\gamma\delta} \omega_{\delta\beta}\right) |\Omega_{\{\alpha,\beta\}.c}|^{-1} \quad (5)$$

where  $c = d \setminus \{\alpha, \beta\}$ ,  $\Omega_{cc}^{-1} = (\Omega_{cc})^{-1}$ , and

$$\Omega_{\{\alpha,\beta\}.c} = \Omega_{\{\alpha,\beta\}\{\alpha,\beta\}} - \Omega_{\{\alpha,\beta\}c} \Omega_{cc}^{-1} \Omega_{c\{\alpha,\beta\}}.$$

Since  $\alpha$  and  $\beta$  are not adjacent in  $(\mathcal{G}_{\leftrightarrow})^a$  there is no edge  $\alpha \leftrightarrow \beta$  in  $\mathcal{G}$ , hence  $\omega_{\alpha\beta} = 0$ . Now consider each term in the sum (5). If there is no edge  $\alpha \leftrightarrow \gamma$  or no edge  $\delta \leftrightarrow \beta$  then  $\omega_{\alpha\gamma} (\Omega_{cc}^{-1})_{\gamma\delta} \omega_{\delta\beta} = 0$ . If there are edges  $\alpha \leftrightarrow \gamma$  and  $\delta \leftrightarrow \beta$  in  $\mathcal{G}$  then  $\gamma \neq \delta$  as otherwise  $\alpha$  and  $\beta$  would be collider connected in  $\mathcal{G}_{\leftrightarrow}$ , and further  $\gamma$  and  $\delta$  are not collider connected in  $(\mathcal{G}_c)_{\leftrightarrow}$ . Hence by the inductive hypothesis,  $(\Omega_{cc}^{-1})_{\gamma\delta} = 0$ . Thus every term in the sum is zero and we are done.  $\square$

An alternative proof follows from the Markov properties of undirected graphical Gaussian models (see Lauritzen, 1996): view the specification of  $\Omega$  formally as if it were an inverse covariance matrix for a model represented by an undirected graph  $\mathcal{U}$ . Then  $\alpha$  and  $\beta$  are not collider connected in  $\mathcal{G}$  if and only if  $\alpha$  and  $\beta$  are not connected in  $\mathcal{U}$ . Hence by the global Markov property

for undirected graphs,  $\alpha$  and  $\beta$  are marginally independent, so  $(\Omega^{-1})_{\alpha\beta} = 0$ . (We thank S. Lauritzen for this observation.)

It also follows directly from the previous Lemma (and this discussion) that  $\Omega^{-1}$  will be block diagonal. (We thank N. Wermuth for this observation.)

**Corollary 8.11** *Let  $\mathcal{G}$  be an ancestral graph with  $\alpha \leftrightarrow \beta$  in  $\mathcal{G}$ . Let  $\mathcal{G}'$  be the subgraph formed by removing the  $\alpha \leftrightarrow \beta$  edge in  $\mathcal{G}$ . If  $\alpha$  and  $\beta$  are not adjacent in  $(\mathcal{G}'_{\leftrightarrow})^a$  then*

$$(\Omega^{-1})_{\alpha\beta} = -\Phi(\alpha \leftrightarrow \beta) |\Omega_{\{\alpha, \beta\}.c}|^{-1},$$

where  $c = d \setminus \{\alpha, \beta\}$ ,  $\Phi$  is a covariance function for  $\mathcal{G}$ , and  $\Omega$  is the associated matrix.

Note that we adopt the convention:  $\Omega_{\{\alpha, \beta\}.c} = \Omega_{\{\alpha, \beta\}}$  when  $c = \emptyset$ .

**Proof:** By the argument used in the proof of Lemma 8.10, it is clear that the sum in equation (5) is equal to 0. The result then follows since, by definition,  $\omega_{\alpha\beta} = \Phi(\alpha \leftrightarrow \beta)$ .  $\square$

## 8.2 Gaussian Independence Models

A Gaussian independence model,  $\mathcal{N}(\mathcal{I})$ , is the set of non-singular normal distributions obeying the independence relations in  $\mathcal{I}$ :

$$\mathcal{N}(\mathcal{I}) \equiv \mathcal{N}_{|V|} \cap \mathcal{P}(\mathcal{I})$$

where  $V$  is the set of vertices in  $\mathcal{I}$ . As noted in Section 7, normal distributions in  $\mathcal{N}(\mathcal{I})$  may also satisfy other independence relations.

**Proposition 8.12** *If  $\mathcal{G}'$  is a subgraph of  $\mathcal{G}$  then  $\mathcal{N}(\mathcal{I}_m(\mathcal{G}')) \subseteq \mathcal{N}(\mathcal{I}_m(\mathcal{G}))$ .*

**Proof:** Follows directly from Proposition 3.12.  $\square$

**Theorem 8.13** *If  $\mathcal{G}_1, \mathcal{G}_2$  are two ancestral graphs then*

$$\mathcal{N}(\mathcal{I}_m(\mathcal{G}_1)) = \mathcal{N}(\mathcal{I}_m(\mathcal{G}_2)) \text{ if and only if } \mathcal{I}_m(\mathcal{G}_1) = \mathcal{I}_m(\mathcal{G}_2).$$

**Proof:** If  $\mathcal{I}_m(\mathcal{G}_1) = \mathcal{I}_m(\mathcal{G}_2)$ , then  $\mathcal{N}(\mathcal{I}_m(\mathcal{G}_1)) = \mathcal{N}(\mathcal{I}_m(\mathcal{G}_2))$  by definition.

By Theorem 7.5 there is a normal distribution  $N_1$  that is faithful to  $\mathcal{I}_m(\mathcal{G}_1)$ . Hence

$$\langle A, B \mid Z \rangle \in \mathcal{I}_m(\mathcal{G}_1) \iff A \perp\!\!\!\perp B \mid Z [N_1].$$

Since  $\mathcal{N}(\mathcal{I}_m(\mathcal{G}_1)) = \mathcal{N}(\mathcal{I}_m(\mathcal{G}_2))$ ,  $N_1 \in \mathcal{N}(\mathcal{I}_m(\mathcal{G}_2))$ , hence  $\mathcal{I}_m(\mathcal{G}_2) \subseteq \mathcal{I}_m(\mathcal{G}_1)$ . The reverse inclusion may be argued symmetrically.  $\square$

### 8.3 Equivalence of Gaussian Parametrizations and Independence Models for Maximal Ancestral Graphs

The main result of this section is the following:

**Theorem 8.14** *If  $\mathcal{G}$  is a maximal ancestral graph then*

$$\mathcal{N}(\mathcal{G}) = \mathcal{N}(\mathfrak{I}_m(\mathcal{G})).$$

In words, if  $\mathcal{G}$  is a maximal ancestral graph then the set of normal distributions that may be obtained by parametrizing  $\mathcal{G}$  is exactly the set of normal distributions that obey the independence relations in  $\mathfrak{I}_m(\mathcal{G})$ .

Note that Wermuth et al. (1994) refer to a ‘parametrization’ of an independence model when describing a parametrization of a (possibly proper) subset of  $\mathcal{N}(\mathfrak{I})$ . To distinguish their usage from the stronger sense in which the term is used here, we may say that a parametrization is *full* if all distributions in  $\mathcal{N}(\mathfrak{I})$  are parametrized. In these terms Theorem 8.14 states that if  $\mathcal{G}$  is maximal then the parametrization of  $\mathcal{G}$  described in Section 8.1 is a full parametrization of  $\mathcal{N}(\mathfrak{I}_m(\mathcal{G}))$ .

#### 8.3.1 $\mathcal{N}(\mathcal{G})$ when $\mathcal{G}$ is not maximal

If  $\mathcal{G}$  is not maximal then  $\mathcal{N}(\mathcal{G})$  is a proper subset of  $\mathcal{N}(\mathfrak{I}_m(\mathcal{G}))$ , as the following example illustrates: consider the non-maximal ancestral graph  $\mathcal{G}$  shown in Figure 9(a). Since  $\mathfrak{I}_m(\mathcal{G}) = \emptyset$ ,  $\mathcal{N}(\mathfrak{I}_m(\mathcal{G})) = \mathcal{N}_4$ , the saturated model. However, there are 10 free parameters in  $\mathcal{N}_4$  and yet there are only 5 edges and 4 vertices, giving 9 parameters in  $\mathcal{N}(\mathcal{G})$ . Direct calculation shows that

$$\sigma_{\gamma\delta} - \frac{\sigma_{\gamma\alpha}\sigma_{\alpha\delta}}{\sigma_{\alpha\alpha}} - \frac{\sigma_{\gamma\beta}\sigma_{\beta\delta}}{\sigma_{\beta\beta}} + \frac{\sigma_{\gamma\alpha}\sigma_{\alpha\beta}\sigma_{\beta\delta}}{\sigma_{\alpha\alpha}\sigma_{\beta\beta}} = 0$$

where  $\sigma_{\phi\psi} = (\Sigma_{\mathcal{G}\Phi})_{\phi\psi}$ . This will clearly not hold for all distributions in  $\mathcal{N}_4$ .

#### 8.3.2 If $\mathcal{G}$ is maximal then $\mathcal{N}(\mathfrak{I}_m(\mathcal{G})) \subseteq \mathcal{N}(\mathcal{G})$

We first require two Lemmas:

**Lemma 8.15** *Let  $\mathcal{G} = (V, E)$  be an ancestral graph,  $\epsilon$  an edge in  $E$  with endpoints  $(\alpha, \beta)$  and  $V = \text{ant}_{\mathcal{G}}(\{\alpha, \beta\})$ . If  $\mathcal{G}' = (V, E \setminus \{\epsilon\})$  is maximal, then for an arbitrary covariance function  $\Phi$  for  $\mathcal{G}$ ,  $(\Sigma_{\mathcal{G}\Phi}^{-1})_{\alpha\beta} = 0$  implies  $\Phi(\epsilon) = 0$ .*

In words, if in a graph  $\mathcal{G}$ , removing an edge,  $\epsilon$ , between  $\alpha$  and  $\beta$  results in a graph that is still maximal, then in any distribution  $N_{\mathcal{G}\mu\Phi}$  obtained from a parametrization  $\langle \mu, \Phi \rangle$  of  $\mathcal{G}$ , if the partial correlation between  $\alpha$  and  $\beta$  given  $V \setminus \{\alpha, \beta\}$  is zero, then  $\Phi$  assigns zero to the edge  $\epsilon$ .

**Proof:** There are three cases, depending on the type of the edge  $\epsilon$ :

(1)  $\epsilon$  is undirected;

In this case  $\alpha, \beta \in \text{un}_{\mathcal{G}}$ . Then by Lemma 8.4,

$$(\Sigma^{-1})_{\alpha\beta} = (\Lambda + B_{du}^{\top} \Omega^{-1} B_{du})_{\alpha\beta}.$$

However, since  $V = \text{ant}_{\mathcal{G}}(\{\alpha, \beta\})$ ,  $d = \emptyset$ , hence  $(\Sigma^{-1})_{\alpha\beta} = (\Lambda)_{\alpha\beta} = \Phi(\alpha - \beta)$ , so  $\Phi(\alpha - \beta) = 0$  as required.

(2)  $\epsilon$  is directed;

Without loss, suppose  $\alpha \leftarrow \beta$ . It now follows from Lemma 8.4, that

$$\begin{aligned} (\Sigma^{-1})_{\alpha\beta} &= B_{d\{\alpha\}}^{\top} \Omega^{-1} B_{d\{\beta\}} \\ &= \sum_{\gamma, \delta \in d} b_{\gamma\alpha} (\Omega^{-1})_{\gamma\delta} b_{\delta\beta}. \end{aligned}$$

Now,  $b_{\gamma\alpha} = 0$  for  $\alpha \neq \gamma$  since  $\text{ch}_{\mathcal{G}}(\alpha) = \emptyset$ , and  $b_{\alpha\alpha} = 1$  by definition. Hence

$$(\Sigma^{-1})_{\alpha\beta} = \sum_{\delta \in d} (\Omega^{-1})_{\alpha\delta} b_{\delta\beta}.$$

Since  $\beta \rightarrow \alpha$ ,  $\beta \in \text{ant}_{\mathcal{G}}(\alpha)$ , so  $V = \text{ant}_{\mathcal{G}}(\alpha)$ . Thus if  $\delta \in V$ ,  $\alpha \neq \delta$ , and  $\alpha$  and  $\delta$  are connected by a path  $\pi$  in  $\mathcal{G}_{\leftrightarrow}$ , containing more than one edge (see p.60), then  $\pi$  is a primitive inducing path between  $\alpha$  and  $\delta$  in  $\mathcal{G}$ . But this is a contradiction, since  $\delta \in \text{ant}_{\mathcal{G}}(\alpha)$ , and yet by Lemma 4.5 (ii),  $\delta \notin \text{ant}_{\mathcal{G}}(\alpha)$ . Hence by Lemma 8.10,  $(\Omega^{-1})_{\alpha\delta} = 0$  for  $\delta \neq \alpha$ . Consequently,

$$(\Sigma^{-1})_{\alpha\beta} = (\Omega^{-1})_{\alpha\alpha} b_{\alpha\beta} = (\Omega^{-1})_{\alpha\alpha} \Phi(\alpha \rightarrow \beta).$$

As  $\Omega$  is positive definite,  $(\Omega^{-1})_{\alpha\alpha} > 0$ , hence  $\Phi(\alpha \rightarrow \beta) = 0$ .



(3)  $\epsilon$  is bi-directed.

Again it follows from Lemma 8.4, that

$$\begin{aligned} (\Sigma^{-1})_{\alpha\beta} &= B_{d\{\alpha\}}^\top \Omega^{-1} B_{d\{\beta\}} \\ &= \sum_{\gamma, \delta \in d} b_{\gamma\alpha} (\Omega^{-1})_{\gamma\delta} b_{\delta\beta}. \end{aligned}$$

As  $\text{ch}_{\mathcal{G}}(\{\alpha, \beta\}) = \emptyset$ ,  $b_{\gamma\alpha} = 0$  for  $\gamma \neq \alpha$ , and likewise  $b_{\delta\beta} = 0$  for  $\delta \neq \beta$ . By definition,  $b_{\beta\beta} = b_{\alpha\alpha} = 1$ . Since, by hypothesis,  $\mathcal{G}'$  is maximal,  $\alpha$  and  $\beta$  are not adjacent in  $(\mathcal{G}'_{\leftrightarrow})^a$ , so

$$(\Sigma^{-1})_{\alpha\beta} = (\Omega^{-1})_{\alpha\beta} = -\Phi(\alpha \leftrightarrow \beta) |\Omega_{\{\alpha, \beta\}.c}|^{-1},$$

the second equality following by Corollary 8.11. Hence  $\Phi(\alpha \leftrightarrow \beta) = 0$  as required.  $\square$

Note that case (2) could alternatively have been proved by direct appeal to the interpretation of  $\Phi(\alpha \leftarrow \beta)$  as a regression coefficient, as shown by Theorem 8.7. However, such a proof is not available in case (3), and we believe that the current proof provides greater insight into the role played by the graphical structure.

The next Lemma provides the inductive step in the proof of the claim which follows.

**Lemma 8.16** *Let  $\mathcal{G} = (V, E)$  be an ancestral graph and  $\epsilon$  an edge in  $E$ . If  $\mathcal{G}' = (V, E \setminus \{\epsilon\})$  is maximal, and  $\text{un}_{\mathcal{G}} = \text{un}_{\mathcal{G}'}$ , then*

$$\mathcal{N}(\mathcal{G}) \cap \mathcal{N}(\mathfrak{I}_m(\mathcal{G}')) \subseteq \mathcal{N}(\mathcal{G}').$$

**Proof:** Let  $N \in \mathcal{N}(\mathcal{G}) \cap \mathcal{N}(\mathfrak{I}_m(\mathcal{G}'))$ , with covariance matrix  $\Sigma$ , and parametrization  $\Phi_{\mathcal{G}}$ . Let  $\epsilon$  have endpoints  $\alpha, \beta$ . Since  $\text{un}_{\mathcal{G}} = \text{un}_{\mathcal{G}'}$  it is sufficient to show that  $\Phi_{\mathcal{G}}(\epsilon) = 0$ , because in this case, the restriction of  $\Phi_{\mathcal{G}}$  to the edges (and vertices) in  $\mathcal{G}'$  is a parametrization of  $\mathcal{G}'$ , hence  $N \in \mathcal{N}(\mathcal{G}')$ .

Let  $A = \text{ant}_{\mathcal{G}'}(\{\alpha, \beta\}) = \text{ant}_{\mathcal{G}}(\{\alpha, \beta\})$ . Since  $\alpha, \beta$  are not adjacent in  $\mathcal{G}'$  and  $\mathcal{G}'$  is maximal, it follows from Corollary 5.3 that

$$\langle \{\alpha\}, \{\beta\} \mid \text{ant}(\{\alpha, \beta\}) \setminus \{\alpha, \beta\} \rangle \in \mathfrak{I}_m(\mathcal{G}').$$

Since  $N \in \mathcal{N}(\mathfrak{I}_m(\mathcal{G}'))$ , it then follows from standard properties of the Normal distribution that  $(\Sigma_{AA}^{-1})_{\alpha\beta} = 0$ . By Lemma 8.5  $\Sigma_{AA}^{-1}$  is parametrized by  $\Phi_A$ , the restriction of  $\Phi_{\mathcal{G}}$  to the edges and vertices in the induced subgraph  $\mathcal{G}_A$ . The result then follows by applying Lemma 8.15 to  $\mathcal{G}_A$ , giving  $\Phi_A(\epsilon) = \Phi_{\mathcal{G}}(\epsilon) = 0$ .  $\square$

We are now in a position to prove that if  $\mathcal{G}$  is maximal then all distributions in  $\mathcal{N}(\mathfrak{I}_m(\mathcal{G}))$  may be obtained by parametrizing  $\mathcal{G}$ . This constitutes one half of Theorem 8.14:

**Claim:** If  $\mathcal{G}$  is maximal then  $\mathcal{N}(\mathfrak{I}_m(\mathcal{G})) \subseteq \mathcal{N}(\mathcal{G})$ .

**Proof:** Suppose  $N \in \mathcal{N}(\mathfrak{I}_m(\mathcal{G}))$ . Let  $\tilde{\mathcal{G}}$  be the completed graph defined in Section 5.2. By Theorem 8.9,  $\mathcal{N}_{|V|} = \mathcal{N}(\tilde{\mathcal{G}})$ , hence  $N \in \mathcal{N}(\tilde{\mathcal{G}})$ . By Theorem 5.6, there exists a sequence of maximal ancestral graphs  $\tilde{\mathcal{G}} \equiv \mathcal{G}_0, \dots, \mathcal{G}_r \equiv \mathcal{G}$  where  $r$  is the number of non-adjacent vertices in  $\mathcal{G}$  and  $\text{un}_{\mathcal{G}_0} = \dots = \text{un}_{\mathcal{G}_r}$ . Now by Proposition 8.12,

$$\mathcal{N}(\mathfrak{I}_m(\mathcal{G}_r)) \subset \dots \subset \mathcal{N}(\mathfrak{I}_m(\mathcal{G}_0)) = \mathcal{N}_{|V|}$$

hence  $N \in \mathcal{N}(\mathfrak{I}_m(\mathcal{G}_i))$ , for  $0 \leq i \leq r$ . We thus may apply Lemma 8.16  $r$ -times to show successively

$$N \in \mathcal{N}(\mathcal{G}_i) \cap \mathcal{N}(\mathfrak{I}_m(\mathcal{G}_{i+1})) \text{ implies } N \in \mathcal{N}(\mathcal{G}_{i+1})$$

for  $i = 0$  to  $r - 1$ . Hence  $N \in \mathcal{N}(\mathcal{G}_r) = \mathcal{N}(\mathcal{G})$  as required.  $\square$

### 8.3.3 $\mathcal{N}(\mathcal{G})$ obeys the global Markov property for $\mathcal{G}$

The following lemma provides a partial converse to Lemma 8.15.

**Lemma 8.17** *If  $\Phi$  is a covariance function for an ancestral graph  $\mathcal{G} = (V, E)$ , and  $\alpha, \beta \in V$  are not adjacent in  $(\mathcal{G})^a$  then  $(\Sigma_{\mathcal{G}\Phi}^{-1})_{\alpha\beta} = 0$ .*

**Proof:** There are two cases to consider:

- (1)  $\alpha \notin \text{un}_{\mathcal{G}}$  or  $\beta \notin \text{un}_{\mathcal{G}}$ ;

By Lemma 8.4:

$$(\Sigma_{\mathcal{G}\Phi}^{-1})_{\alpha\beta} = \sum_{\gamma, \delta \in d} b_{\gamma\alpha}(\Omega^{-1})_{\gamma\delta} b_{\delta\beta} \quad (6)$$

If  $b_{\gamma\alpha} \neq 0$  and  $b_{\delta\beta} \neq 0$  then there are edges  $\alpha \rightarrow \gamma$ ,  $\beta \rightarrow \delta$  in  $\mathcal{G}$ , hence  $\gamma \neq \delta$ ,  $\beta \neq \gamma$  and  $\alpha \neq \delta$  since otherwise  $\alpha$  and  $\beta$  are adjacent in  $(\mathcal{G})^a$ . Further, there is no path between  $\gamma$  and  $\delta$  in  $\mathcal{G}_{\leftrightarrow}$  since if there were,  $\alpha$  and  $\beta$  would be collider connected in  $\mathcal{G}$ , hence adjacent in  $(\mathcal{G})^a$ . Thus  $\gamma$  and  $\delta$  are not adjacent in  $(\mathcal{G}_{\leftrightarrow})^a$  and so by Lemma 8.10  $(\Omega^{-1})_{\gamma\delta} = 0$ . Consequently every term in the sum in (6) is zero as required.

(2)  $\alpha, \beta \in \text{un}_{\mathcal{G}}$ .

Again by Lemma 8.4:

$$(\Sigma_{\mathcal{G}\Phi}^{-1})_{\alpha\beta} = \lambda_{\alpha\beta} + \sum_{\gamma, \delta \in d} b_{\gamma\alpha}(\Omega^{-1})_{\gamma\delta}b_{\delta\beta}. \quad (7)$$

If  $\alpha, \beta$  are not adjacent in  $(\mathcal{G})^a$  then  $\alpha$  and  $\beta$  are not adjacent in  $\mathcal{G}$ . Hence  $\lambda_{\alpha\beta} = 0$ . The argument used in case (1) may now be repeated to show that every term in the sum in (7) is zero.  $\square$

The next lemma proves the second half of Theorem 8.14. It does not require  $\mathcal{G}$  to be maximal, so we state it as a separate lemma.

**Lemma 8.18** *If  $\mathcal{G}$  is an ancestral graph then  $\mathcal{N}(\mathcal{G}) \subseteq \mathcal{N}(\mathfrak{I}_m(\mathcal{G}))$ .*

In words, any normal distribution obtained by parametrizing an ancestral graph  $\mathcal{G}$  obeys the global Markov property for  $\mathcal{G}$ .

**Proof:** Suppose that  $\langle X, Y \mid Z \rangle \in \mathfrak{I}_m(\mathcal{G})$ . If  $\nu \in \text{ant}_{\mathcal{G}}(X \cup Y \cup Z) \setminus (X \cup Y \cup Z)$  then in  $(\mathcal{G}_{\text{ant}(X \cup Y \cup Z)})^a$  either  $\nu$  is separated from  $X$  by  $Z$ , or from  $Y$  by  $Z$ . Hence  $X$  and  $Y$  may always be extended to  $X^*, Y^*$  respectively, such that  $\langle X^*, Y^* \mid Z \rangle \in \mathfrak{I}_m(\mathcal{G})$  and  $X^* \cup Y^* \cup Z = \text{ant}_{\mathcal{G}}(X \cup Y \cup Z)$ . Since the multivariate normal density is strictly positive, for an arbitrary  $N \in \mathcal{N}_{[V]}$ ,

$$(C5) \quad A \perp\!\!\!\perp B \mid C \cup D \text{ and } A \perp\!\!\!\perp C \mid B \cup D \text{ implies } A \perp\!\!\!\perp B \cup C \mid D$$

(see Dawid, 1980). By repeated application of C5 it is sufficient to show that for each pair  $\alpha, \beta$  with  $\alpha \in X^*$ ,  $\beta \in Y^*$ ,

$$\alpha \perp\!\!\!\perp \beta \mid (Z \cup X^* \cup Y^*) \setminus \{\alpha, \beta\} \quad [N],$$

or equivalently  $(\Sigma_{AA}^{-1})_{\alpha\beta} = 0$ , where  $A = X^* \cup Y^* \cup Z$ . Since  $\langle X^*, Y^* \mid Z \rangle \in \mathfrak{I}_m(\mathcal{G})$ ,  $\alpha$  and  $\beta$  are not adjacent in  $(\mathcal{G}_A)^a$ . The result then follows from Lemma 8.5 and Lemma 8.17.  $\square$

Lemmas 8.17 and 8.18 are based on Lemma 3.1.6 and Theorem 3.1.8 in Koster (1999a), though these results concern a different class of graphs (see Section 9.2). An alternative proof of Lemma 8.18 for ancestral graphs without undirected edges is given in Spirtes et al. (1996, 1998).

### 8.3.4 Distributional equivalence of Markov equivalent models

The following corollary states that two maximal ancestral graphs are Markov equivalent if and only if the corresponding Gaussian models are equivalent.

**Corollary 8.19** *For maximal ancestral graphs,  $\mathcal{G}_1, \mathcal{G}_2$ ,*

$$\mathfrak{I}_m(\mathcal{G}_1) = \mathfrak{I}_m(\mathcal{G}_2) \text{ if and only if } \mathcal{N}(\mathcal{G}_1) = \mathcal{N}(\mathcal{G}_2).$$

**Proof:**

$$\begin{aligned} \mathcal{N}(\mathcal{G}_1) = \mathcal{N}(\mathcal{G}_2) &\Leftrightarrow \mathcal{N}(\mathfrak{I}_m(\mathcal{G}_1)) = \mathcal{N}(\mathfrak{I}_m(\mathcal{G}_2)) && \text{by Theorem 8.14;} \\ &\Leftrightarrow \mathfrak{I}_m(\mathcal{G}_1) = \mathfrak{I}_m(\mathcal{G}_2) && \text{by Theorem 8.13.} \end{aligned}$$

□

**Corollary 8.20** *If  $\mathcal{G} = (V, E)$  is an ancestral graph and  $S \cup L \subset V$ ,*

$$\text{if } N \in \mathcal{N}(\mathcal{G}) \text{ then } N_{x_L}^{x_S=x_S} \in \mathcal{N}(\mathcal{G}_L^S).$$

*for all  $x_S \in \mathbb{R}^{|S|}$ .*

**Proof:** By Lemma 8.18  $\mathcal{N}(\mathcal{G}) \subseteq \mathcal{N}(\mathfrak{I}_m(\mathcal{G}))$ . Hence by normality and Theorem 7.1  $N_{x_L}^{x_S=x_S} \in \mathcal{N}(\mathfrak{I}_m(\mathcal{G})_L^S)$ . Finally, by Corollary 4.19  $\mathcal{G}_L^S$  is maximal, hence

$$\mathcal{N}(\mathfrak{I}_m(\mathcal{G})_L^S) = \mathcal{N}(\mathfrak{I}_m(\mathcal{G}_L^S)) = \mathcal{N}(\mathcal{G}_L^S),$$

by Theorem 4.18 and Theorem 8.14. □

Suppose that we postulate a Gaussian model  $\mathcal{N}(\mathcal{G})$  with complex structure, such as a DAG containing latent variables and/or selection variables. This corollary is significant because it guarantees that if  $\mathcal{N}(\mathcal{G})$  contains the ‘true’ distribution  $N^*$ , and we then simplify  $\mathcal{G}$  to a model for the observed variables,  $\mathcal{N}(\mathcal{G}_L^S)$ , then the new model will contain the true ‘observable’ distribution obtained by marginalizing the unobserved variables and conditioning on the selection variables,  $N_{x_L}^{x_S=x_S}$ . (The distribution  $N_{x_L}^{x_S=x_S}$  is termed ‘observable’ because it is the distribution over the observed variables  $(V \setminus (S \cup L))$  in the ‘selected’ subpopulation for which  $X_S = x_S$ . In general this will obviously *not* be the distribution observed in a finite sample.)

## 8.4 Gaussian Ancestral Graph Models Are Curved Exponential Families

Let  $\mathcal{S}$  be a full regular exponential family of dimension  $m$  with natural parameter space  $\Theta \subseteq \mathbb{R}^m$ , so  $\mathcal{S} = \{P_\theta \mid \theta \in \Theta\}$ . If  $U$  is an open neighbourhood in  $\Theta$ , then  $\mathcal{S}^U = \{P_\theta \mid \theta \in U\}$ . Let  $\mathcal{S}_0$  be a sub-family of  $\mathcal{S}$ , with  $\Theta_0$  the corresponding subset of  $\Theta$ .

If  $A$  is open in  $\mathbb{R}^m$  then a function  $f : A \rightarrow \mathbb{R}^m$  is a *diffeomorphism of  $A$  onto  $f(A)$*  if  $f(\cdot)$  is one-to-one, smooth (infinitely differentiable), and of full rank everywhere on  $A$ . Corollary A.3 in Kass and Vos (1997) states that a function  $f$  is a diffeomorphism if it is smooth, one-to-one, and the inverse  $f^{-1} : f(A) \rightarrow A$  is also smooth.

Theorem 4.2.1 in Kass and Vos (1997) states that a subfamily  $\mathcal{S}_0$  of an  $m$ -dimensional regular exponential family  $\mathcal{S}$  is a locally parametrized curved exponential family of dimension  $k$  if for each  $\theta_0 \in \Theta_0$  there is an open neighbourhood  $U$  in  $\Theta$  containing  $\theta_0$  and a diffeomorphism  $f : U \rightarrow \mathbb{R}^k \times \mathbb{R}^{m-k}$ , and

$$\mathcal{S}_0^U = \{P_\theta \in \mathcal{S}^U \mid f(\theta) = (\psi, 0)\}.$$

We use the following fact in the next Lemma:

**Proposition 8.21** *If  $f$  is a rational function defined everywhere on a set  $D$  then  $f^{(n)}$  is a rational function defined everywhere on  $D$ .*

**Proof:** The proof is by induction on  $n$ . Suppose  $f^{(n)} = g_n/h_n$ , where  $g_n, h_n$  are polynomials, and  $h_n > 0$  on  $D$ . Then  $f^{(n+1)} = (h_n g'_n - g_n h'_n)/h_n^2$  from which the conclusion follows (since  $h_n^2 > 0$  on  $D$ ).  $\square$

Let  $\Sigma_{|V|}^+$  denote the cone of positive definite  $|V| \times |V|$  matrices.

**Lemma 8.22** *If  $\mathcal{G}$  is a complete ancestral graph then the mapping*

$$f_{\mathcal{G}} : \Phi(\mathcal{G}) \rightarrow \mathbb{R}^{|V|} \times \Sigma_{|V|}^+ \quad \text{given by} \quad \langle \mu, \Phi \rangle \mapsto \langle \mu, \Sigma_{\mathcal{G}\Phi} \rangle$$

*is a diffeomorphism from  $\Phi(\mathcal{G})$  to  $\mathbb{R}^{|V|} \times \Sigma_{|V|}^+$ .*

**Proof:** Corollary 8.8 establishes that  $f_{\mathcal{G}}$  is one-to-one. Further, by Theorem 8.9,  $\mathcal{N}(\mathcal{G}) = \mathcal{N}_{|V|}$  hence

$$f(\Phi(\mathcal{G})) = \mathbb{R}^{|V|} \times \Sigma_{|V|}^+.$$

It remains to show that  $f_{\mathcal{G}}, f_{\mathcal{G}}^{-1}$  are smooth. It follows from equation (1) that the components of  $f_{\mathcal{G}}$  are rational functions of  $\langle \mu, \Phi \rangle$ , defined for all  $\langle \mu, \Phi \rangle \in \Phi(\mathcal{G})$ . Hence, by Proposition 8.21,  $f_{\mathcal{G}}$  is smooth. Similarly, equation (2) establishes that  $f_{\mathcal{G}}^{-1}$  is smooth.  $\square$

**Theorem 8.23** *For an ancestral graph  $\mathcal{G}(V, E)$ ,  $\mathcal{N}(\mathcal{G})$  is a curved exponential family, with dimension  $2 \cdot |V| + |E|$ .*

**Proof:** This follows from the definition of  $\mathcal{N}(\mathcal{G})$ , the existence of a complete ancestral supergraph of  $\mathcal{G}$  (Lemma 5.5), Lemma 8.22 and Theorem 4.2.1 of Kass and Vos (1997), referred to above.  $\square$

The BIC criterion for the model  $\mathcal{N}(\mathcal{G})$  is given by

$$\text{BIC}(\mathcal{G}) = -2 \ln L_{\mathcal{G}}(\hat{\theta}) + \ln(n)(2 \cdot |V| + |E|),$$

where  $n$  is the sample size,  $L_{\mathcal{G}}(\cdot)$  is the likelihood function and  $\hat{\theta}$  is the corresponding MLE for  $\mathcal{N}(\mathcal{G})$ . A consequence of Theorem 8.23 is that  $\text{BIC}(\cdot)$  is an asymptotically consistent criterion for selecting among Gaussian ancestral graph models (see Haughton, 1988).

By contrast, Geiger et al. (2001) have shown that simple discrete DAG models with latent variables do not form curved exponential families.

## 8.5 Parametrization via Recursive Equations with Correlated Errors

The Gaussian model  $\mathcal{N}(\mathcal{G})$  can alternatively be parametrized in two pieces via the factorization of the density:

$$f(x_V) = f(x_{\text{un}_{\mathcal{G}}})f(x_{V \setminus \text{un}_{\mathcal{G}}} \mid x_{\text{un}_{\mathcal{G}}}) \quad (8)$$

The undirected component  $f_{\text{un}_{\mathcal{G}}}$  may be parametrized via an undirected graphical Gaussian model also known as a covariance selection model (see Lauritzen, 1996, Dempster, 1972).

The directed component,  $f(x_{V \setminus \text{un}_{\mathcal{G}}} \mid x_{\text{un}_{\mathcal{G}}})$ , may be parametrized via a set of recursive equations as follows:

- (i) Associate with each  $\nu$  in  $V \setminus \text{un}_{\mathcal{G}}$  a linear equation, expressing  $X_{\nu}$  as a linear function of the variables for the parents of  $\nu$  plus an error term:

$$X_{\nu} = \mu_{\nu} + \sum_{\pi \in \text{pa}(\nu)} b_{\nu\pi}^* \cdot X_{\pi} + \epsilon_{\nu}.$$

- (ii) Specify a non-singular multivariate normal distribution over the error variables  $(\epsilon_{\nu})_{\nu \in V \setminus \text{un}_{\mathcal{G}}}$  (with mean zero) satisfying the condition that

$$\text{if there is no edge } \alpha \leftrightarrow \beta \text{ in } \mathcal{G}, \text{ then } \text{Cov}(\epsilon_{\alpha}, \epsilon_{\beta}) = 0,$$

but otherwise unrestricted.

Note that  $b_{\alpha\beta}^* = -b_{\alpha\beta}$  under the parametrization specified in Section 8.1. The conditional distribution,  $f(x_{V \setminus \text{un}_{\mathcal{G}}} \mid x_{\text{un}_{\mathcal{G}}})$ , is thus parametrized via a simultaneous equation model, of the kind used in econometrics and psychometrics since the 1940's. We describe the system as 'recursive' because the equations may be arranged in upper triangular form, possibly with correlated errors. (Note that some authors only use this term if, in addition, the errors are uncorrelated.) As shown in Theorem 8.7 the set of recursive equations described here also has the special property that the linear coefficients may be consistently estimated via regression of each variable on its parents. This does not hold for recursive equations in general.

### 8.5.1 Estimation procedures

The parametrization described above thus breaks  $\mathcal{N}(\mathcal{G})$  into an undirected graphical Gaussian model and a set of recursive equations with correlated errors. This result is important for the purposes of statistical inference because software packages exist for estimating these models: MIM (Edwards, 1995) fits undirected Gaussian models via the IPS algorithm; AMOS (Arbuckle, 1997), EQS (Bentler, 1986), Proc CALIS (SAS Publishing, 1995) and LISREL (Jöreskog and Sörbom, 1995) are packages which fit structural equation models via numerical optimization. Fitting the two components separately is possible in view of the factorization of the likelihood given by equation (8) and the variation independence of the parameters in these pieces (see Proposition 8.2).

It should be noted that the equations used in the parametrization above are a very special (and simple) subclass of the much more general class of models that structural equation modelling packages can fit e.g. they only contain observed variables. This motivates the future development of special purpose fitting procedures.

### 8.5.2 Path diagrams

Path diagrams, introduced by Sewall Wright (1921, 1934), contain directed and bi-directed edges, but no undirected edges, and are used to represent structural equations in exactly the way described in (i) and (ii) above. Hence we have the following:

**Proposition 8.24** *If  $\mathcal{G}$  is an ancestral graph containing no undirected edges then  $\mathcal{N}(\mathcal{G})$  is the model obtained by regarding  $\mathcal{G}$  as a path diagram.*

Further results relating path diagrams and graphical models are described in Spirtes et al. (1998), Koster (1999a,b, 1996) and Spirtes (1995). The relationship between Gaussian ancestral graph models and Seemingly Unrelated Regression (SUR) models (Zellner, 1962) is discussed in Richardson et al. (1999).

## 8.6 Canonical DAGs Do Not Provide a Full Parametrization

It was proved in Section 6 that the canonical DAG  $\mathcal{D}(\mathcal{G})$  provides a way of reducing the global Markov property for ancestral graphs to that of DAGs. It is thus natural to consider whether the associated Gaussian independence model could be parametrized via the usual parametrization of this DAG. In fact, this does not parametrize all distributions in  $\mathcal{N}(\mathfrak{I}_m(\mathcal{G}))$  as shown in the following example:

Consider the ancestral graph  $\mathcal{G}_1$ , and the associated canonical DAG,  $\mathcal{D}(\mathcal{G}_1)$  shown in Figure 17(i-a) and (i-b). Since  $\mathfrak{I}_m(\mathcal{G}_1) = \emptyset$ ,  $\mathcal{N}(\mathfrak{I}_m(\mathcal{G}_1)) = \mathcal{N}_3$  the saturated model on 3 variables. However, if  $N$  is a distribution given by a parametrization of  $\mathcal{D}(\mathcal{G}_1)$ , then it follows by direct calculation that

$$\min\{\rho_{ab}, \rho_{bc}, \rho_{ac}\} < \frac{1}{\sqrt{2}}$$



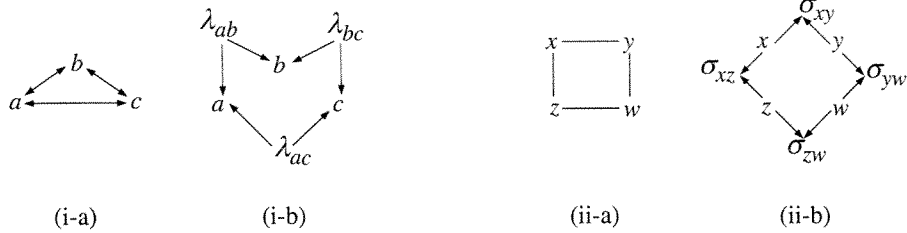


Figure 17: (i-a) An ancestral graph  $\mathcal{G}_1$ ; (i-b) the corresponding canonical DAG,  $\mathcal{D}(\mathcal{G}_1)$ ; (ii-a) an ancestral graph  $\mathcal{G}_2$ ; (ii-b) the canonical DAG,  $\mathcal{D}(\mathcal{G}_2)$ .

where  $\rho_{vw}$  is the correlation between  $X_v$  and  $X_w$  (see Spirtes et al., 1998). Since this does not hold for all distributions in  $\mathcal{N}_3$ , there are normal distributions  $N \in \mathcal{N}(\mathcal{I}_m(\mathcal{G}_1))$  for which there is no distribution  $N^* \in \mathcal{N}(\mathcal{D}(\mathcal{G}_1))$  such that  $N = N^*_{\{\lambda_{ab}, \lambda_{bc}, \lambda_{ac}\}}$ .

Lauritzen (1998) p.12 gives an analogous example for conditioning, by considering the graph  $\mathcal{G}_2$ , with canonical DAG,  $\mathcal{D}(\mathcal{G}_2)$ , shown in Figure 17(ii-a) and (ii-b). Lauritzen shows that there are normal distributions  $N \in \mathcal{N}(\mathcal{I}_m(\mathcal{G}_2))$ , for which there is no distribution  $N^* \in \mathcal{N}(\mathcal{D}(\mathcal{G}_2))$  such that  $N = N^*_{\{\sigma_{xy}, \sigma_{xz}, \sigma_{zw}, \sigma_{yw}\}}$ .

These negative results are perhaps surprising given the very simple nature of the structure in  $\mathcal{D}(\mathcal{G})$ , but serve to illustrate the complexity of the sets of distributions represented by such models.

## 9 Relation to Other Work

The problem of constructing graphical representations for the independence structure of DAGs under marginalizing *and* conditioning was originally posed by N. Wermuth in 1994 in a lecture at CMU. Wermuth, Cox and Pearl developed an approach to this problem based on *summary graphs* (see Wermuth et al., 1994, 1999, Cox and Wermuth, 1996, Wermuth and Cox, 2000). More recently J. Koster has introduced another class of graphs, called *MC-graphs*, together with an operation of marginalizing and conditioning. (See Koster, 2000, Koster, 1999a, Koster, 1999b.)

In Figure 18 we show two examples of data generating processes, together with the maximal ancestral graph, summary graph and MC-graphs resulting

after marginalizing (i) and conditioning (ii).

Simple representations for DAGs under marginalization alone were proposed by Verma (1993), who defined an operation of *projection* which transforms a DAG with latent variables to another DAG in which each latent variable has exactly two children both of which are observed (called a ‘semi-Markovian model’). The operation is defined so that the DAG and its projection are Markov equivalent over the common set of observed variables. This approach does not lead to a full parametrization of the independence model for the reasons discussed in Section 8.6.

In this section we will briefly describe the classes of summary graphs and MC-graphs. We then outline the main differences and similarities to the class of maximal ancestral graphs. Finally we discuss the relation between ancestral graphs and chain graphs.

## 9.1 Summary Graphs

A *summary graph* is a graph containing three types of edge  $\rightarrow$ ,  $-$ ,  $----$ . Directed cycles may not occur in a summary graph, but it is possible for there to be a dashed line ( $\alpha ---- \beta$ ) and at the same time a directed path from  $\alpha$  to  $\beta$ . Thus there may be two edges between a pair of vertices, e.g.  $\alpha \rightrightarrows \beta$ . This is the only combination of multiple edges that is permitted. The separation criterion for summary graphs is equivalent to m-separation after substituting bi-directed edges ( $\leftrightarrow$ ) for dashed edges ( $----$ ).

Wermuth et al. (1999) presents an algorithm for transforming a summary graph so as to represent the independence structure remaining among the variables after marginalizing and conditioning. This procedure will not, in general, produce a graph that obeys a pairwise Markov property, hence there may be a pair of vertices  $\alpha, \beta$  that are not adjacent and yet there is no subset  $Z$  of the remaining vertices for which the model implies  $\alpha \perp\!\!\!\perp \beta \mid Z$ . The graph in Figure 18(i-c) illustrates this. There is no edge between  $a$  and  $c$ , and yet  $a \not\perp\!\!\!\perp c$  and  $a \not\perp\!\!\!\perp c \mid b$ . This example also illustrates that there may be more edges than pairs of adjacent vertices in a summary graph.

Wermuth and Cox (2000) present a new method for constructing a summary graph based on applying ‘sweep’ operators to matrices whose entries indicate the presence or absence of edges. Kauermann (1996) analyses the subset of summary graphs that only involve dashed edges, which are also known as covariance graphs.

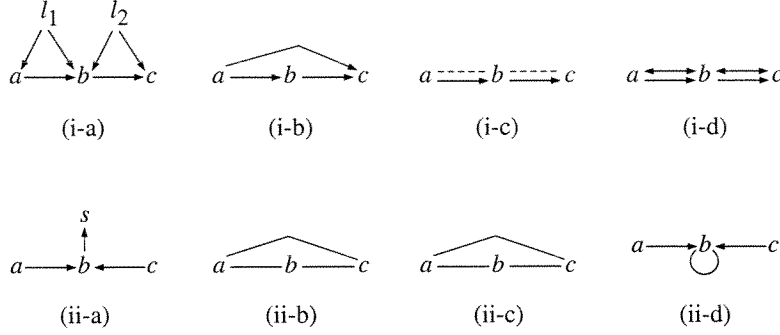


Figure 18: (i-a) a DAG generating process  $\mathcal{D}_1$ ; (i-b) the ancestral graph  $\mathcal{D}_1^{\emptyset}_{\{l_1, l_2\}}$ ; the summary graph (i-c) and MC-graph (i-d) resulting from marginalizing  $l_1, l_2$  in  $\mathcal{D}_1$ . (ii-a) a DAG generating process  $\mathcal{D}_2$ ; (ii-b) the ancestral graph  $\mathcal{D}_2^{\{s\}}_{\emptyset}$ ; the summary graph (ii-c) and MC-graph (ii-d) resulting from conditioning on  $s$  in  $\mathcal{D}_2$ .

## 9.2 MC-Graphs

Koster (1999a,b) considers *MC-graphs*, which include the three edge types  $\rightarrow, \leftrightarrow, \leftarrow$ , but in addition may also contain undirected self-loops (see vertex  $b$  in Figure 18(ii-d)). Up to four edges may be present between a pair of vertices, e.g.

$$\alpha \begin{matrix} \overleftarrow{\phantom{\rightarrow}} \\ \overleftrightarrow{\phantom{\rightarrow}} \\ \overrightarrow{\phantom{\rightarrow}} \end{matrix} \beta.$$

The global Markov property used for MC-graphs is identical to the m-separation criterion (Koster names the criterion ‘d-separation’ because it is a natural generalization of the criterion for DAGs). Koster presents a procedure for transforming the graph under marginalizing and conditioning. As with the summary graph procedure the transformed graph will not generally obey a pairwise Markov property, and may have more edges than there are pairs of vertices.

## 9.3 Comparison of Approaches

The three classes of graphs: ancestral graphs, summary graphs and MC-graphs have been developed with similar goals in mind, hence it is not sur-

prising that in certain respects they are similar. However, there are also a number of differences between the approaches.

For the rest of this section we will ignore the notational distinction between dashed lines (---) and bi-directed edges ( $\leftrightarrow$ ) by treating them as if they were the same symbol.

### 9.3.1 Graphical and Markov structure

The following (strict) inclusions relate the classes of graphs:

$$\text{maximal ancestral} \subset \text{ancestral} \subset \text{summary} \subset \text{MC}.$$

Essentially the same separation criterion is used for ancestral graphs, summary graphs and MC-graphs. Further, defining  $\mathbb{I}[\cdot]$  to denote a class of independence models, we have:

$$\mathbb{I}[\text{maximal ancestral}] = \mathbb{I}[\text{ancestral}] = \mathbb{I}[\text{summary}] \subset \mathbb{I}[\text{MC}].$$

The first equality is Theorem 5.1, the second equality follows by a construction similar to the canonical DAG (Section 6). The last inclusion is strict because MC-graphs include directed cyclic graphs which, in general, are not Markov equivalent to any DAG under marginalization and conditioning (Richardson, 1996). In addition, there are MC-graphs which cannot be obtained by applying the marginalizing and conditioning transformation to a graph containing only directed edges: Figure 19 gives an example. Thus the class of MC-graphs is larger than required for representing directed graphs under marginalizing and conditioning. The direct analogues to Theorems 6.3 and 6.4 do not hold.

In the summary graph formed by the procedures described in Wermuth et al. (1999), Wermuth and Cox (2000), the configurations  $-\gamma \text{---}$  and  $-\gamma \leftarrow$  never occur. This is equivalent to condition (ii) in the definition of an ancestral graph. Consequently, as noted by Wermuth et al. (1999) a decomposition of the type shown in Figure 4 is possible for summary graphs. However, though directed cycles do not occur in summary graphs, the analogue to condition (i) does not hold, since it is possible to have an edge  $\alpha \text{---} \beta$  and a directed path from  $\alpha$  to  $\beta$ .

The marginalizing and conditioning transformation operations for summary graphs and MC-graphs are ‘local’ in that they make changes to triples of adjacent vertices. In contrast the transformation  $\mathcal{G} \mapsto \mathcal{G}_L^S$  requires pairwise

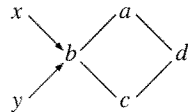


Figure 19: An MC-graph which cannot be obtained by applying the marginalizing and conditioning transformation given by Koster (2000) to a graph which contains only directed edges. Further, the independence model corresponding to this MC-graph cannot be obtained by marginalizing and conditioning an independence model represented by a directed graph.

tests of m-separation to be carried out in order to determine the adjacencies present in  $\mathcal{G}_L^s$ . This may make the transformation harder for a human to carry out. On the other hand the transformation given by Wermuth is recursive, and tests for the existence of an m-connecting path can be performed by a recursive procedure that only examines triples of adjacent vertices. It can be said that the MC-graph and summary graph transformations may in general be performed in fewer steps than the ancestral graph transformation.

However, a price is paid for not performing these tests of m-separation: whereas  $\mathcal{G}_L^s$  always obeys a pairwise Markov property (Corollary 4.19), the summary graphs and MC graphs resulting from the transformations do not do so in general. This is a disadvantage in a visual representation of an independence model insofar as it conflicts with the intuition, based on separation in undirected graphs, that if two vertices are not connected by an edge then they are not *directly* connected and hence may be made independent by conditioning on an appropriate subset of the other vertices.

### 9.3.2 Gaussian parametrization

For summary graphs, as for ancestral graphs, the Gaussian parametrization consists of a conditional distribution and a marginal distribution. Once again, the marginal parametrization is specified via a covariance selection model and the conditional distribution via a system of structural equations of the type used in econometrics and psychometrics as described in Section 8.5 (see Cox and Wermuth, 1996). Under this parametrization one parameter is associated with each edge and vertex in the graph.

As described above, it is possible for a summary graph to contain more

edges than there are pairs of adjacent vertices. Consequently, the Gaussian model associated with a summary graph will not be identified in general, and the analogous result to Corollary 8.8 will not hold. Thus the summary graph model will sometimes contain more parameters than needed to parametrize the corresponding Gaussian independence model.

On the other hand, as mentioned in the previous section, summary graphs do not satisfy a pairwise Markov property, and hence the associated model will not parametrize all Gaussian distributions satisfying the Markov property for the graph. In particular, the comments concerning non-maximal ancestral graphs apply to summary graphs (see Section 8.3.1). In other words, parametrization of a summary graph does not, in general, lead to a full parametrization of the independence model (see Theorem 8.14). In this sense the summary graph model sometimes contains too few parameters.

As a consequence, two Markov equivalent summary graphs may represent different sets of Gaussian distributions, so the analogue to Corollary 8.19 does not hold. Thus for the purpose of parametrizing Gaussian independence models, the class of maximal ancestral graphs has advantages over summary graphs (and non-maximal ancestral graphs).

It should be stressed, however, that the fact that a summary graph model may impose additional non-Markovian restrictions can be seen as an advantage insofar as it may lead to more parsimonious models. For this purpose ideally one would wish to develop a graphical criterion that would also allow the non-independence restrictions to be read from the graph. In addition, one would need to show that the analogue to Corollary 8.20 held for the transformation operation, so that any non-Markovian restrictions imposed by the model associated with the transformed summary graph were also imposed by the original model. Otherwise there is the possibility that while the original model contained the true population distribution, by introducing an additional non-Markovian constraint, the model after transformation no longer contains the true distribution. The approach in Wermuth and Cox (2000) considers the parametrization as derived from the original DAG in the manner of structural equation models with latent variables. Under this scheme the same summary graph may have different parametrizations. An advantage of this scheme is that the strengths of the associations may be calculated if we know the parameters of the generating DAG.

Finally, note that the linear coefficients occurring in the equations in a summary graph model do not always have a population interpretation as regression coefficients. This is because there may be an edge  $\alpha \cdots \beta$  and a

directed path from  $\alpha$  to  $\beta$ . (However, coefficients associated with edges  $v \rightarrow \delta$  where  $v$  is a vertex in the undirected subgraph do have this interpretation, as noted by Wermuth and Cox (2000).) Hence the analogue to Theorem 8.7 does not hold for all summary graphs.

Koster (1999a,b) does not discuss parameterization of MC-graphs, however all of the above comments will apply to any parametrization which associates one parameter with each vertex and edge. Indeed, under such a scheme identifiability will be more problematic than for summary graphs because MC-graphs permit more edges between vertices in addition to self-loops.

## 9.4 Chain Graphs

A mixed graph containing no partially directed cycles, and no bi-directed edges is called a *chain graph*. (Recall that a partially directed cycle is an anterior path from  $\alpha$  to  $\beta$ , together with an edge  $\beta \rightarrow \alpha$ .) There is an extensive body of work on chain graphs. (See Lauritzen (1996) for a review.)

As was shown in Lemma 3.2(c) an ancestral graph does not contain partially directed cycles, hence we have the following:

**Proposition 9.1** *If  $\mathcal{G}$  is an ancestral graph containing no bi-directed edges then  $\mathcal{G}$  is a chain graph.*

In fact, it is easy to see that the set of ancestral chain graphs are the recursive ‘causal’ graphs introduced by Kiiveri et al. (1984); see also Lauritzen and Richardson (2002) and Richardson (2001).

Two different global Markov properties have been proposed for chain graphs. Lauritzen and Wermuth (1989) and Frydenberg (1990a) proposed the first Markov property for chain graphs. More recently Andersson et al. (2001, 1996) have proposed an alternative Markov property. We will denote the resulting independence models  $\mathcal{I}_{LWF}(\mathcal{G})$  and  $\mathcal{I}_{AMP}(\mathcal{G})$  respectively.

The m-separation criterion as applied to chain graphs produces yet another Markov property. (This observation is also made by Koster (1999a).) In general all three properties will be different, as illustrated by the chain graph in Figure 20(i). Under both the AMP and LWF properties  $a \perp\!\!\!\perp b$  in  $CG_1$ , but this does not hold under m-separation because the path  $a \rightarrow x - y \leftarrow b$  m-connects  $a$  and  $b$  given the empty set. The AMP property implies  $a \perp\!\!\!\perp y$ , while this is not implied by m-separation or the LWF property. Note that

under m-separation this chain graph is Markov equivalent to an undirected graph.

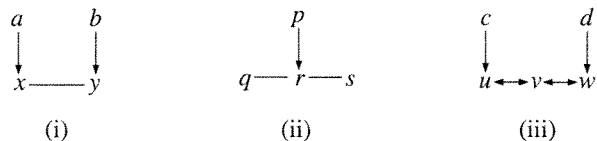


Figure 20: Chain graphs that are not Markov equivalent to any ancestral graph under (i) the LWF property, (ii) the AMP property; (iii) an ancestral graph for which there is no Markov equivalent chain graph (under either Markov property).

However, if we restrict our attention to ancestral graphs then we have the following proposition:

**Proposition 9.2** *If  $\mathcal{G}$  is an ancestral graph which is also a chain graph then*

$$\mathfrak{I}_m(\mathcal{G}) = \mathfrak{I}_{LWF}(\mathcal{G}) = \mathfrak{I}_{AMP}(\mathcal{G}).$$

This proposition is an immediate consequence of clause (i) in the definition of an ancestral graph which implies that there are no immoralities, flags or bi-flags in an ancestral mixed graph. (See Frydenberg (1990a) and Andersson et al. (1996) for the relevant definitions.)

Finally, note that under both the LWF and AMP Markov properties there exist chain graphs that are not Markov equivalent to any ancestral graph. Examples are shown in Figure 20(i),(ii). It follows that these Markov models could not have arisen from any DAG generating process. (See Lauritzen and Richardson (2002) and Richardson (1998) for further discussion.) Conversely, Figure 20(iii) shows an example of an independence model represented by an ancestral graph that is not Markov equivalent to any chain graph (under either chain graph Markov property).

## 10 Discussion

In this paper we have introduced the class of ancestral graph Markov models. The purpose in introducing this class was to be able to characterize the



Markov structure of a DAG model under marginalizing and conditioning. To this end we defined a graphical transformation  $\mathcal{G} \mapsto \mathcal{G}_{[L]}^S$ , which corresponded to marginalizing and conditioning the corresponding independence model (Theorem 4.18).

If a DAG model containing latent or selection variables is hypothesized as the generating mechanism for a given system then this transformation will allow a simple representation of the Markov model induced on the observed variables.

However, often graphical models are used for exploratory data analysis, where little is known about the generating structure. In such situations the existence of this transformation provides a guarantee: *if* the data were generated by an unknown DAG containing hidden variables then we are ensured that there exists an ancestral graph which can represent the resulting Markov structure over the observed variables. Hence the problem of additional and misleading edges encountered in the introduction may be avoided. In this context the transformation provides a justification for using the class of ancestral graphs.

However, any interpretation of the types of edge present in an ancestral graph which was arrived at via an exploratory analysis should take into account that there may exist (many) different graphs that are Markov equivalent. Spirtes and Richardson (1997) present a polynomial-time algorithm for testing Markov equivalence of two ancestral graphs. Spirtes et al. (1995, 1999) describe an algorithm for inferring structural features that are common to all maximal ancestral graphs in a Markov equivalence class. For instance, there are Markov equivalence classes in which every member contains a directed path from some vertex  $\alpha$  to a second vertex  $\beta$ ; likewise in other Markov equivalence classes no member contains a directed path from  $\alpha$  to  $\beta$ . At the time of writing there is not yet a full characterization of common features, such as exists for DAG Markov equivalence classes (see Andersson et al., 1997, Meek, 1995a).

Finally, we showed that maximal ancestral graphs lead to a natural parametrization of the set of Gaussian distributions obeying the global Markov property for the graph. Conditions for the existence and uniqueness of maximum likelihood estimates for these models is currently an open question.

Development of a parametrization for discrete distributions is another area of current research. Richardson (2002) describes a local Markov property for a class of graphs that includes all ancestral graphs without undirected edges. This local Markov property is equivalent to the global Markov prop-

erty, and may thus facilitate the development of a discrete parametrization.

## Acknowledgments

The work presented here has benefited greatly from the frequent interaction that the authors have had with Nanny Wermuth and Jan Koster during the last five years. In particular, as mentioned, the idea of studying independence models under conditioning was first raised by Wermuth, as was the notion of a canonical DAG. The proof techniques employed by Koster in lecture notes originally prepared for a course at the Fields Institute (Koster, 1999a) had a considerable influence on the development of Sections 7 and 8. Much of this interaction took place at workshops that were part of the European Science Foundation program on Highly Structure Stochastic Systems.

The authors are also grateful to Ayesha Ali, Steen Andersson, Heiko Bailer, Moulinath Banerjee, Sir David Cox, Clark Glymour, David Heckerman, Steffen Lauritzen, David Madigan, Chris Meek, Michael Perlman, Jamie Robins, Tamas Rudas, Richard Scheines, Jim Q. Smith, Milan Studený, Larry Wasserman and Jacob Wegelin for stimulating conversations and helpful comments on earlier drafts.

This research was supported by the U.S. National Science Foundation (DMS-9972008 & DMS-9873442), the U.S. Office of Naval Research and the Isaac Newton Institute, Cambridge, UK where the first author was a Rosenbaum fellow from July-December 1997.

## A Appendix

### A.1 Definition of a Mixed Graph

Let  $\mathcal{E} = \{ -, \leftarrow, \rightarrow, \leftrightarrow \}$ , be the set of edges. Let  $\mathfrak{P}(\mathcal{E})$  denote the power set of  $\mathcal{E}$ . Formally, a *mixed graph*  $\mathcal{G} = (V, E)$  is an ordered pair consisting of a finite set  $V$ , and a mapping  $E : V \times V \rightarrow \mathfrak{P}(\mathcal{E})$ , subject to the following restrictions:

$$\begin{aligned} E(\alpha, \alpha) &= \emptyset \\ - &\in E(\alpha, \beta) \iff - \in E(\beta, \alpha) \\ \leftarrow &\in E(\alpha, \beta) \iff \rightarrow \in E(\beta, \alpha) \\ \leftrightarrow &\in E(\alpha, \beta) \iff \leftrightarrow \in E(\beta, \alpha). \end{aligned}$$

The *induced subgraph*,  $\mathcal{G}_A$  of  $\mathcal{G}$  on  $A \subseteq V$ , is  $(A, E|_A)$  where  $E|_A$  is the natural restriction of  $E$  to  $A \times A$ .

## References

- Andersson, S., D. Madigan, M. Perlman, and C. Triggs (1995). On the relation between conditional independence models determined by finite distributive lattices and by directed acyclic graphs. *Journal of Statistical Planning and Inference* 48, 25–46.
- Andersson, S., D. Madigan, M. Perlman, and C. Triggs (1997). A graphical characterization of lattice conditional independence models. *Annals of Mathematics and Artificial Intelligence* 21, 27–50.
- Andersson, S. and M. Perlman (1998). Normal linear regression models with recursive graphical Markov structure. *Journal of Multivariate Analysis* 66, 133–187.
- Andersson, S. A., D. Madigan, and M. D. Perlman (1996). An alternative Markov property for chain graphs. In F. V. Jensen and E. Horvitz (Eds.), *Uncertainty in Artificial Intelligence: Proceedings of the 12<sup>th</sup> Conference*, San Francisco, pp. 40–48. Morgan Kaufmann.
- Andersson, S. A., D. Madigan, and M. D. Perlman (1997). A characterization of Markov equivalence classes for acyclic digraphs. *Ann. Statist.* 25, 505–541.
- Andersson, S. A., D. Madigan, and M. D. Perlman (2001). Alternative Markov properties for chain graphs. *Scand. J. Statist.* 28(1), 33–86.
- Arbuckle, J. L. (1997). *AMOS User's Guide Version 3.6*. Chicago: SPSS.
- Bentler, P. M. (1986). *Theory and implementation of EQS: A structural equations program*. Los Angeles: BMDP Statistical Software.
- Cooper, G. F. (1995). Causal discovery from data in the presence of selection bias. In D. Fisher (Ed.), *Preliminary papers of the Fifth International Workshop on AI and Statistics, January 4-7, Fort Lauderdale, Florida*, pp. 140–150.

- Cox, D. R. and N. Wermuth (1996). *Multivariate Dependencies: Models, Analysis and Interpretation*. London: Chapman and Hall.
- Darroch, J., S. Lauritzen, and T. Speed (1980). Markov fields and log-linear interaction models for contingency tables. *Ann. Statist.* 8, 522–539.
- Dawid, A. (1980). Conditional independence for statistical operations. *Ann. Statist.* 8, 598–617.
- Dempster, A. P. (1972). Covariance selection. *Biometrics* 28, 157–175.
- Edwards, D. M. (1995). *Introduction to Graphical Modelling*. New York: Springer-Verlag.
- Frydenberg, M. (1990a). The chain graph Markov property. *Scand. J. Statist.* 17, 333–353.
- Frydenberg, M. (1990b). Marginalization and collapsibility in graphical interaction models. *Ann. Statist.* 18, 790–805.
- Geiger, D. (1990). *Graphoids: a qualitative framework for probabilistic inference*. Ph. D. thesis, UCLA.
- Geiger, D., D. Heckerman, H. King, and C. Meek (2001). Stratified exponential families: Graphical models and model selection. *Ann. Statist.* 29, 505–529.
- Geiger, D. and J. Pearl (1990). On the logic of causal models. In R. D. Shachter, T. S. Levitt, L. N. Kanal, and J. F. Lemmer (Eds.), *Uncertainty in Artificial Intelligence IV*, Amsterdam, pp. 136–147. North-Holland.
- Haughton, D. (1988). On the choice of a model to fit the data from an exponential family. *Ann. Statist.* 16, 342–355.
- Jöreskog, K. and D. Sörbom (1995). *LISREL 8: User's Reference Guide*. Chicago: Scientific Software International.
- Kass, R. E. and P. W. Vos (1997). *Geometrical foundations of asymptotic inference*. New York: Wiley.
- Kauermann, G. (1996). On a dualization of graphical Gaussian models. *Scand. J. Statist.* 23, 105–116.

- Kiiveri, H., T. Speed, and J. Carlin (1984). Recursive causal models. *Journal of the Australian Mathematical Society, Ser. A* 36, 30–52.
- Koster, J. T. A. (1996). Markov properties of non-recursive causal models. *Ann. Statist.* 24, 2148–2178.
- Koster, J. T. A. (1999a, September). Linear structural equations and graphical models. Lecture notes, The Fields Institute, Toronto.
- Koster, J. T. A. (1999b, October). On the validity of the Markov interpretation of path diagrams of Gaussian structural equation systems with correlated errors. *Scand. J. Statist.* 26(3), 413–431.
- Koster, J. T. A. (2000). Marginalizing and conditioning in graphical models. Technical report, Erasmus University, Rotterdam.
- Lauritzen, S. (1996). *Graphical Models*. Number 17 in Oxford Statistical Science Series. Oxford, UK: Clarendon Press.
- Lauritzen, S. L. (1979). Lectures on contingency tables. Technical report, Institute of Mathematical Statistics, University of Copenhagen.
- Lauritzen, S. L. (1998). Generating mixed hierarchical interaction models by selection. Technical Report R-98-2009, Department of Mathematics, University of Aalborg.
- Lauritzen, S. L. and T. S. Richardson (2002). Chain graph models and their causal interpretations (with discussion). *J. Roy. Statist. Soc. Ser. B* 64, ?–?
- Lauritzen, S. L. and N. Wermuth (1989). Graphical models for association between variables, some of which are qualitative and some quantitative. *Ann. Statist.* 17, 31–57.
- Meek, C. (1995a). Causal inference and causal explanation with background knowledge. In P. Besnard and S. Hanks (Eds.), *Uncertainty in Artificial Intelligence: Proceedings of the 11<sup>th</sup> Conference*, San Francisco, pp. 403–410. Morgan Kaufmann.

- Meek, C. (1995b). Strong completeness and faithfulness in Bayesian networks. In P. Besnard and S. Hanks (Eds.), *Uncertainty in Artificial Intelligence: Proceedings of the 11<sup>th</sup> Conference*, San Francisco, pp. 411–418. Morgan Kaufmann.
- Pearl, J. (1988). *Probabilistic Reasoning in Intelligent Systems*. Morgan Kaufmann.
- Richardson, T. S. (1996). *Models of feedback: interpretation and discovery*. Ph. D. thesis, Carnegie-Mellon University.
- Richardson, T. S. (1998). Chain graphs and symmetric associations. In M. Jordan (Ed.), *Learning in Graphical Models*, Number 89 in NATO Science Series D, pp. 231–260. Kluwer.
- Richardson, T. S. (2001). Chain graphs which are maximal ancestral graphs are recursive causal graphs. Technical Report 387, Department of Statistics, University of Washington.
- Richardson, T. S. (2002). Markov properties for acyclic directed mixed graphs. *Scand. J. Statist.* 29, ?–?
- Richardson, T. S., H. Bailer, and M. Banerjee (1999). Tractable structure search in the presence of latent variables. In D. Heckerman and J. Whittaker (Eds.), *Preliminary papers of the Seventh International Workshop on AI and Statistics, January 4-7, Fort Lauderdale, Florida*, pp. 142–151. Morgan Kaufmann.
- Robins, J. (1997). Causal inference from complex longitudinal data. In M. Berkane (Ed.), *Latent variable modelling and applications to causality*, Number 120 in Lecture notes in statistics, pp. 69–117. New York: Springer-Verlag.
- SAS Publishing (1995). *SAS/STAT User's Guide, Version 6* (Fourth ed.). Cary, NC: SAS Publishing.
- Settimi, R. and J. Q. Smith (1998). On the geometry of Bayesian graphical models with hidden variables. In G. Cooper and S. Moral (Eds.), *Uncertainty in Artificial Intelligence: Proceedings of the 14<sup>th</sup> Conference*, San Francisco, pp. 472–479. Morgan Kaufmann.

- Settimi, R. and J. Q. Smith (1999). Geometry, moments and Bayesian networks with hidden variables. In D. Heckerman and J. Whittaker (Eds.), *Preliminary papers of the Seventh International Workshop on AI and Statistics, January 4-7, Fort Lauderdale, Florida*, pp. 293-298.
- Spirtes, P. (1995). Directed cyclic graphical representations of feedback models. In P. Besnard and S. Hanks (Eds.), *Uncertainty in Artificial Intelligence: Proceedings of the 11<sup>th</sup> Conference*, San Francisco, pp. 491-498. Morgan Kaufmann.
- Spirtes, P., C. Glymour, and R. Scheines (1993). *Causation, Prediction and Search*. Number 81 in Lecture Notes in Statistics. Springer-Verlag.
- Spirtes, P., C. Meek, and T. S. Richardson (1995). Causal inference in the presence of latent variables and selection bias. In P. Besnard and S. Hanks (Eds.), *Uncertainty in Artificial Intelligence: Proceedings of the 11<sup>th</sup> Conference*, San Francisco, pp. 403-410. Morgan Kaufmann.
- Spirtes, P., C. Meek, and T. S. Richardson (1999). Causal inference in the presence of latent variables and selection bias. In C. Glymour and G. F. Cooper (Eds.), *Computation, Causality and Discovery*, pp. 211-252. Cambridge, Mass.: MIT Press.
- Spirtes, P. and T. S. Richardson (1997). A polynomial-time algorithm for determining DAG equivalence in the presence of latent variables and selection bias. In D. Madigan and P. Smyth (Eds.), *Preliminary papers of the Sixth International Workshop on AI and Statistics, January 4-7, Fort Lauderdale, Florida*, pp. 489-501.
- Spirtes, P., T. S. Richardson, C. Meek, R. Scheines, and C. Glymour (1996). Using d-separation to calculate zero partial correlations in linear models with correlated errors. Technical Report CMU-PHIL-72, Department of Philosophy, Carnegie-Mellon University.
- Spirtes, P., T. S. Richardson, C. Meek, R. Scheines, and C. Glymour (1998). Using path diagrams as a structural equation modelling tool. *Sociological Methods and Research* 27, 182-225.
- Verma, T. (1993). Graphical aspects of causal models. Technical Report R-191, Cognitive Systems Laboratory, UCLA.

- Verma, T. and J. Pearl (1990). Equivalence and synthesis of causal models. In M. Henrion, R. Shachter, L. Kanal, and J. Lemmer (Eds.), *Uncertainty in Artificial Intelligence: Proceedings of the 6<sup>th</sup> Conference*, Mountain View, CA, pp. 220–227. Association for Uncertainty in AI.
- Verma, T. and J. Pearl (1991). Equivalence and synthesis of causal models. Technical Report R-150, Cognitive Systems Laboratory, UCLA.
- Wermuth, N. and D. Cox (2000). A sweep operator for triangular matrices and its statistical applications. Technical Report 00-04, ZUMA Institute, Mannheim, Germany.
- Wermuth, N., D. Cox, and J. Pearl (1994). Explanations for multivariate structures derived from univariate recursive regressions. Technical Report 94-1, University of Mainz, Mainz, Germany.
- Wermuth, N., D. Cox, and J. Pearl (1999). Explanations for multivariate structures derived from univariate recursive regressions. Technical Report Revision of 94-1, University of Mainz, Mainz, Germany.
- Whittaker, J. (1990). *Graphical Models in applied multivariate statistics*. Chichester: John Wiley.
- Wright, S. (1921). Correlation and Causation. *J. Agricultural Research* 20, 557–585.
- Wright, S. (1934). The method of path coefficients. *Ann. Math. Statist.* 5, 161–215.
- Zellner, A. (1962). An efficient method of estimating seemingly unrelated regression equations and tests for aggregation bias. *JASA* 57, 348–368.